

THIS WEEK

EDITORIALS

ZOOLOGY Online-only species naming made official at last **p.178**

WORLD VIEW The corruption of open-access scientific publishing **p.179**



COOLING Horses sweat, people perspire and model houses glow **p.180**

Count on me

Sometimes, the use of metrics to assess the value of scientists is unavoidable. So let's come up with the best measure possible.

In an ideal world, scientists applying for grants or jobs would be judged holistically — balancing quantitative measures such as their publication record against indications of their potential from recommendation letters, personal interactions and other activities. So even if a candidate had not generated many papers, it would count in their favour if the few they had published had received positive post-publication review (comments, tweets and blog posts, for instance). Also favourable would be a tendency to ask insightful questions at talks that lead to valuable discussions and new experiments, or a willingness to share reagents and expertise with their colleagues. That would be ideal. But that is not the world in which most scientists live.

Instead, hiring committees and grant reviewers sweat through hundreds of applications, often with only enough time to give each submission a cursory glance. In 2010, a *Nature* poll found that most administrators say that metrics — quantifiable measures of scientists' achievements — matter less in job decisions than scientists often think (see *Nature* **465**, 860–862; 2010), but good peer review is often simply not possible.

As a result, evaluators are increasingly turning to metrics, such as total citation count and the *h*-index, a measure of both the quality and quantity of papers (a scientist has an *h*-index of 12 if they have published 12 papers that have each received at least 12 citations). Naturally, many scientists object to such cold quantification of their contribution. Plus, all metrics have obvious flaws — a paper may gather many citations not because of its importance, but because it is in a large field that publishes frequently, so generates more opportunities for citations. Review articles, which may not add much to the research, count the same as original research papers, which contribute a great deal. And all existing metrics capture only what a scientist has done, not what he or she might be capable of. Clearly, there is a need for more and better measures.

On page 201, Daniel Acuna, Stefano Allesina and Konrad Kording suggest an alternative: the future *h*-index. Unlike other metrics, this index estimates a scientist's publication prowess five years or so into the future — a useful timescale for tenure decisions.

Using publicly available data on the history of publication, citation and funding for thousands of neuroscientists, researchers working on the fruitfly *Drosophila* and evolutionary biologists, the authors constructed an algorithm that converts information on a typical scientist's CV — the number of journals published in and articles in top journals for instance — into a number that represents their probable *h*-index in the years that follow.

Outraged? Please send complaints to the usual address. Interested? Calculate your own future *h*-index here: go.nature.com/z4rrroc.

Nature receives thousands of submissions a year, some of which point out the flaws in existing metrics and propose alternatives. We accepted the piece by Acuna *et al.* after submitting it to peer

review. The reviewers and our editors felt that the authors had used appropriate methods to obtain their algorithm, and its predictive values seemed realistic. Furthermore, the authors are cautious about its value, pointing out that it is probably less accurate for scientists in other disciplines, and should not be considered a replacement for peer review. At the very least, the future *h*-index should help to address some problems with the current *h*-index, which tends to favour established scientists because they have had

more time to accrue citations. A forward-looking metric may give a leg up to promising, early-career scientists who don't yet have impressive CVs.

“There is no substitute for examining the research itself to appreciate its value.”

Nevertheless, no one wants their career potential to be reduced to a number. *Nature* publishes many scientific gems that nevertheless achieve few citations; there is no

substitute for examining the research itself to appreciate its value. We know that the idea of a new metric published in these pages will raise some anxieties, and a few hackles. But metrics are already being used, so it is important that they create the most accurate picture possible of someone's potential. Plus, they do hold some advantages over peer review, by helping to eliminate the unconscious biases that can creep into personal evaluations.

In that vein, scientists should continue to hunt for metrics that capture a scientist's true value, including aspects such as teaching, reviewing and public-speaking ability, as well as online responses to publications in blogs and comments — ‘alt-metrics’. We may not live in an ideal world, but we can still improve the recruitment, reward and opportunities for scientists. ■

Secret weapons

US military furtiveness is hindering progress and the development of technology.

In the 1940s, with the Second World War in full swing, Japanese scientists sketched out a plan to build a microwave weapon to shoot down enemy bombers. That idea, perhaps the earliest description of an electromagnetic bomb, encapsulates much of what military officials still hope to achieve with such weapons: disabling electronics (or, in some cases, people) using a powerful energy beam, without causing any collateral physical damage. The US military's attempts to make a practical weapon based on this idea have so far resulted in only one system — at least as far as it has revealed publicly. The Air Force

has built the Active Denial System, a non-lethal high-power microwave weapon supposedly able to deter an angry mob by creating the sensation of being burned.

For decades, the US military has conducted much of its research on such weapons in secret. It has often hinted that it is on the verge of a breakthrough, yet high-power microwave weapons are noticeably absent from modern battlefields and scenes of civil unrest. The military, for the most part, won't discuss its progress — or lack thereof — citing secrecy in the name of national security.

There is nothing unique about the classification of this research: nuclear weapons, stealth aircraft and satellite reconnaissance systems were all developed in secret. Although such furtiveness can legitimately protect US weapons and capabilities, it can also prevent much-needed dissemination of scientific research. And it has all too often concealed a lack of progress.

As we discuss on page 198, this has been the problem with the programme to develop high-power microwave weapons: the little information that has been released points to obvious scientific and technological problems. Crucially, power sources for such devices are often too unwieldy to use. More than ten years after the Active Denial System was first revealed to the public, its size and complexity mean that it is still nearly impossible to deploy. The military rejected the system for use at checkpoints in Iraq because it would have taken 16 hours to cool down the weapon's pulse generator to superconducting temperatures to fire it.

Many records related to the Active Denial System remain classified and inaccessible to the public and the scientific community. The US Air Force's unwillingness to reveal the full scope of its research into the biological effects of high-power microwaves in the 1990s, which included work on their auditory and lethal effects, flies in the face of the defence department's claims that it is interested in classifying only weapons technology, and not science. If, as the Air Force says, the biological research never led to weapons, then there is no reason not to release it.

Work on high-power microwaves designed to take out electronics has not fared much better. Advocates can always claim that classified

programmes are yielding great progress, but information in the public sphere does not paint a rosy picture. Military officials and academics acknowledge that developing compact power sources remains the biggest hurdle. The Air Force and a contractor have touted efforts to develop a high-power microwave cruise missile, but neither will release details that might allow independent experts to judge the programme's potential. The Pentagon is staying quiet on a system developed to take out improvised explosive devices, but what little information is available indicates that — like the Active Denial System — it has proved too cumbersome to use effectively.

"The government must be willing to share data and findings between military labs and academia."

This is not to say that all government spending on high-power microwaves is a waste. Academic funding under Multi-disciplinary University Research Initiatives is contributing to a host of peer-reviewed publications and collaborative research. But for the government to take full advantage of that research, it must be willing to share data and findings between military labs and academia. The defence department's own science board has found that reluctance to share is a barrier to progress.

Getting to the truth about high-power microwaves requires transparency. Independent experts must be able to scrutinize technology to enable scientific-military cooperation and to provide a reality check for those who make fantastic claims about a weapon's potential.

By the time it cancelled the Airborne Laser programme earlier this year, the US defence department had poured billions of dollars into the weapon: a chemical laser in the nose of an aircraft, designed to shoot down ballistic missiles. In the end, the question was not whether the laser would work, but whether it would be usable, given the scientific and technological practicalities of integrating such a complex system. "There's nobody in uniform that I know who believes that this is a workable concept," concluded former US defence secretary Robert Gates, when he finally moved to kill the project. The same concerns would probably be expressed about high-power microwaves — if more information about them were available. ■

The name game

After several years of wrangling, zoologists can now name new species online only.

With access to the Internet, the (official) world of animals and plants will soon be at your fingertips. In a landmark ruling, zoologists last week agreed that newly identified species can be named in online-only publications. Previously, the first official description of anything that crawled, flew, wriggled, walked or swam across Earth needed to be formally written up and recorded in print, where it would remain in perpetuity for future scientists to reference.

That made sense when Henry Fairfield Osborn described *Tyrannosaurus rex* in 1905; less so when Rob Gay named a new theropod, *Kayentavenator elysiae*, in 2010, which helps to explain why Gay broke with convention and claimed the first description of the species in a self-published print-on-demand book.

As technology blurred the distinction between what is published and what is not, some predicted online anarchy, with 'taxonomic vandals' taking to the Internet to self-publish reports of new species. An obvious solution to the problem would have been to extend the rules from print to cover online scientific journals, and to draw the line there. But there were concerns about whether online journals would endure. In a messy compromise, online journals that published descriptions of new species printed and bound several dozen copies of the paper — in case a

twenty-second-century palaeontologist should call. In an even messier compromise, some scientists printed papers from journal websites and posted them to libraries themselves.

No more. The International Commission on Zoological Nomenclature (ICZN), which sets the rules for the naming of new species, announced on 4 September that it was relaxing its code to encompass publication in online-only publications. The change, which followed a vote of 23 in favour to 3 against, with one abstention, comes into force at the start of next year. The amendment allows for descriptions of new species in "widely accessible electronic copies with fixed content and layout". New animal species will also need to be registered with ZooBank.org, the official registry of the ICZN.

It is a sensible move, and one that most in the field should welcome. It comes a year after the International Botanical Congress endorsed online-only publication for new types of plant. In an Editorial at the time (*Nature* 475, 424; 2011), which called on zoologists to follow suit, *Nature* said: "At this point, it seems that there is little reason to continue to demand paper on a shelf to make a species name official."

On hearing the news from the ICZN, one member of *Nature's* staff quipped: "Now to name a dinosaur you don't have to behave like one." But that is a little unfair. Proper taxonomy and a robust archive are crucial to science, and the zoologists were right to consider with care the possible negative aspects of such a change, as well as listening to the clamour to embrace the new. True, the change has been a long time coming. It is overdue, even. Still, when you have been dead and waiting for a name since the Mesozoic era, what are a few extra years? ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunq



Predatory publishers are corrupting open access

Journals that exploit the author-pays model damage scholarly publishing and promote unethical behaviour by scientists, argues Jeffrey Beall.

When e-mail first became available, it was a great innovation that made communication fast and cheap. Then came spam — and suddenly, the innovation wasn't so great. It meant having to filter out irrelevant, deceptive and sometimes offensive messages. It still does.

The same corruption of a great idea is now occurring with scholarly open-access publishing.

Early experiments with open-access publishing, such as the *Journal of Medical Internet Research* and BioMed Central, were very promising. Set up more than a decade ago, they helped to inspire a social movement that has changed academic publishing for the better, lowered costs and expanded worldwide access to the latest research.

Then came predatory publishers, which publish counterfeit journals to exploit the open-access model in which the author pays. These predatory publishers are dishonest and lack transparency. They aim to dupe researchers, especially those inexperienced in scholarly communication. They set up websites that closely resemble those of legitimate online publishers, and publish journals of questionable and downright low quality. Many purport to be headquartered in the United States, United Kingdom, Canada or Australia but really hail from Pakistan, India or Nigeria.

Some predatory publishers spam researchers, soliciting manuscripts but failing to mention the required author fee. Later, after the paper is accepted and published, the authors are invoiced for the fees, typically US\$1,800. Because the scientists are often asked to sign over their copyright to the work as part of the submission process (against the spirit of open access) they feel unable to withdraw the paper and send it elsewhere.

I monitor predatory publishers on my blog, Scholarly Open Access, which has become a forum in which scientists can raise their concerns over the practice. They send me hundreds of e-mails passing on spam solicitations or asking whether a particular publisher is legitimate.

I also get e-mails from the predators' victims. Some have been named as members of editorial boards without their knowledge or permission. Others have had an article partially or completely plagiarized in a predatory journal. Many ask me for advice on where to publish or how to withdraw an article that they wish they hadn't submitted. As a librarian, I do my best to answer the questions I receive, but they often require expertise in the author's field of study. So it is important that more scientists are made aware of the problem.

The predatory publishers and journals often have lofty titles that make them seem legitimate in a list of publications on a CV. Scholarly publishing's traditional role of vetting the best

research is disappearing. Now there is a journal willing to accept almost every article, as long as the author is willing to pay the fee. Authors, rather than libraries, are the customers of open-access publishers, so a powerful incentive to maintain quality has been removed.

Perhaps nowhere are these abuses more acute than in India, where new predatory publishers or journals emerge each week. They are appearing because of the market need — hundreds of thousands of scientists in India and its neighbouring countries need to get published to earn tenure and promotion.

Here, the problem is not just with the publishers. Scientists themselves are also to blame. Many are taking unethical shortcuts and paying for the publication of plagiarized or self-plagiarized work.

Honest scientists stand to lose the most in this unethical quagmire.

When a researcher's work is published alongside articles that are plagiarized, that report on conclusions gained from unsound methodologies or that contain altered photographic figures, it becomes tainted by association. Unethical scientists gaming the system are earning tenure and promotion at the expense of the honest.

The competition for author fees among fraudulent publishers is a serious threat to the future of science communication. To compete in a crowded market, legitimate open-access publishers are being forced to promise shorter submission-to-publication times; this weakens the peer-review process, which takes time to do properly.

To tackle the problem, scholars must resist the temptation to publish quickly and easily. The research community needs to use scholarly social networks such as Connotea and Mendeley

to identify and share information on publishers that deceive, lack transparency or otherwise fail to follow industry standards. Scientific literacy must include the ability to recognize publishing fraud, and libraries must remove predatory publishers from their online catalogues. The worst offenders can usually be discovered without too much effort: their websites are littered with grammatical errors and they list bogus contact details. The borderline cases are more difficult to spot — here, we need open-access zealots to open their eyes to the growing quality problems.

Conventional scholarly publishers have had an important role in validating research, yet too often advocates of open access seem to overlook the importance of validation in online publishing. They promote access at the expense of quality: a shortcoming that tacitly condones the publication of unworthy scientific research. ■

Jeffrey Beall is Scholarly Initiatives Librarian at the University of Colorado Denver.
e-mail: jeffrey.beall@ucdenver.edu

SCIENTIFIC LITERACY
**MUST
INCLUDE
THE ABILITY TO
RECOGNIZE
PUBLISHING
FRAUD.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/bnuu1y

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

ASTRONOMY

On the origin of Ia supernovae

Using a computer model, astronomers in Japan present evidence to support one theory about the origin of type Ia supernovae, which are of predictable luminosity and are therefore useful for estimating distances to remote galaxies. These supernovae are generated when a compact star called a carbon-oxygen white dwarf siphons off a critical amount of matter from a companion star. However, astronomers have yet to directly observe a companion star.

Izumi Hachisu at the University of Tokyo and his team offer a solution to this conundrum. Their model, which accounts for the rotation of white dwarfs, shows that many companion stars have evolved to a state too faint to be seen. The model is consistent with observations of supernova 2011fe.

Astrophys. J. 756, L4 (2012)

MATERIALS

Hydrogel makes buildings sweat

Just as mammals can reduce their body temperature by sweating, so a coating of heat-sensitive hydrogel could 'sweat' to cool buildings.

Wendelin Stark and his colleagues at the Swiss Federal Institute of Technology in Zurich produced a 3-millimetre-thick layer of



thermoresponsive hydrogel. When heated to roughly 32 °C, the gel undergoes a phase transition from a wet state to a dry state, and releases water. With this water could go much of a building's heat.

Miniature model houses (pictured) with roofs coated with the substance were up to 20 °C cooler than uncoated models when exposed to simulated tropical midday Sun. The authors estimate that this translates into annual savings of 220 kilowatt-hours of energy for a detached house. A brief 'rain' recharged the hydrogel with water.

Adv. Mater. <http://dx.doi.org/10.1002/adma.201202574> (2012)

ANIMAL BEHAVIOUR

Seals see glowing prey

Female southern elephant seals are thought to locate their food in deep, dark waters by detecting the bioluminescence of their prey. The seals see best at the wavelength of light produced by bioluminescent organisms.

Jade Vacqu  -Garc  a at the Chiz   Centre for Biological Studies in Villiers-en-Bois, part of the French National Centre for Scientific Research, and her team tested the idea by studying four

female southern elephant seals (*Mirounga leonina*; pictured) living on the Kerguelen islands in the southern Indian Ocean. The researchers used satellite tracking to follow the animals as they made a total of 3,386 dives, and a light sensor to measure bioluminescence. During the dives, the seals' foraging intensity was positively related to the number of bioluminescence events. *PLoS ONE* 7, e43565 (2012)



MOLECULAR BIOLOGY

Red-blood-cell regulator

Human genome-wide association studies have indicated that a common single nucleotide sequence variant influences the size and number of red blood cells. A team led by Harvey Lodish of the Whitehead Institute, and Eric Lander of the Broad Institute, both in Cambridge, Massachusetts, found that this variant in the non-coding sequence reduces the expression of the nearby gene *CCND3*, which is involved in controlling the cell cycle.

Reducing the expression of

the gene in human and mouse red-blood-cell precursors caused the cells to go through fewer divisions, resulting in fewer, but bigger, red blood cells.

Genes Dev. <http://dx.doi.org/10.1101/gad.197020.112> (2012)

MICROBIOLOGY

Populations cooperate

Populations of naturally co-occurring bacteria work together, an analysis suggests.

To test whether bacteria aggregate for reasons beyond access to resources such as food, Martin Polz at the

Massachusetts Institute of Technology in Cambridge and his team created a network map of interactions between 185 strains of *Vibrionaceae* bacteria. They found that antagonistic interactions were much less likely between strains that are as closely related as those within naturally occurring populations.

The authors suggest that the populations act as cohesive units that can defend themselves against other groups of bacteria. The coordination seems to have arisen through horizontal gene transfer, in which population members capable of producing antibiotics passed on antibiotic-resistance genes to their neighbours.

Science 337, 1228–1231 (2012)

MICROBIOLOGY

Ecosystems compete

Bacterial species seem to compete more than they cooperate.

Kevin Foster at the University of Oxford, UK, and Thomas Bell at Imperial College London tested the theory that different bacterial species can form cooperative 'superorganisms'. The team measured the productivity — in terms of CO₂ given off through respiration — of a range of multi-species microbial cultures. After both one week and one month, species in the mixtures were less productive than they would have been if grown separately, suggesting that social living comes at a cost. *Curr. Biol.* <http://dx.doi.org/10.1016/j.cub.2012.08.005> (2012)

NEUROSCIENCE

A treatable form of autism

A rare form of autism that is accompanied by epilepsy and intellectual disability could be treatable with dietary supplements.

Joseph Gleeson and Gaia

Novarino at the University of California, San Diego, and their team identified three affected families that carry mutated versions of a gene called *BCKDK*. The gene encodes an enzyme that slows the breakdown of branched-chain amino acids such as leucine and valine. Mice that lack both copies of the gene had low serum and brain levels of these amino acids, and showed neurological abnormalities and stunted growth. Enriching the animals' diets in branched-chain amino acids reversed these deficits.

Similar dietary enrichment normalized serum levels of branched-chain amino acids in people affected by the *BCKDK* mutation.

Science <http://dx.doi.org/10.1126/science.1224631> (2012)

CONSERVATION BIOLOGY

Tigers and people can coexist

As the world becomes more crowded, setting aside time, as well as space, for other species could help humans to live harmoniously alongside wildlife.

Jianguo Liu of Michigan State University in East Lansing and his team came up with the idea after analysing footage from motion-activated cameras they had set up in and around Nepal's Chitwan National Park. The team detected high levels of both human and tiger activity in the park, and found that the geographical activity patterns of the two species overlapped: any given camera was likely to collect images of both. Yet there was little conflict between humans and tigers in the park, probably because 80% of tiger detections occurred at night, when humans are generally at home. Outside the park, where humans are even thicker on the ground, 95% of tiger detections were at night. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1210490109> (2012)

COMMUNITY CHOICE

The most viewed papers in science

CLIMATE SCIENCE

Calm Sun promotes chilly winters

HIGHLY READ
on www.agu.org
27 Aug–2 Sept

Unusually cold central European winters — such as those of the past two years — could be due to atmospheric circulation anomalies caused by weak solar activity.

Frank Sirocko of the Johannes Gutenberg University of Mainz, Germany, and his team compared historical records of freezing of the river Rhine with a time series of observed sunspot activity. The researchers found that 10 of the 14 major freezing events that have occurred since 1780 took place in or around years with minimal numbers of sunspots, when solar activity was weak.

The effects of weak solar activity may propagate to the lowest layer of Earth's atmosphere, leading to changes in pressure systems over the North Atlantic. These changes would favour the sustained flow of chilly Arctic air to central Europe during winter months, the team suggests.

Geophys. Res. Lett. <http://dx.doi.org/10.1029/2012GL052412> (2012)



PLANT BIOTECHNOLOGY

Engineered plants can use phosphite

Many crops depend on phosphorus-based fertilizers and are threatened by herbicide-resistant weeds. Genetically modified plants that can digest an alternative phosphorus source that normally inhibits plant growth could solve both problems.

Damar López-Arredondo and Luis Herrera-Estrella at the National Polytechnic Institute's Centre for Research and Advanced Studies in Irapuato, Mexico, engineered the model plant *Arabidopsis*, and tobacco plants, to metabolize phosphite, in addition to the orthophosphate

found in standard fertilizer. When phosphite was available, the transgenic plants needed 30–50% less phosphorus to generate the same amount of biomass as that produced in the presence of orthophosphate. The researchers also tested the transgenic plants against weeds. In the presence of orthophosphate, weeds dominated the transgenic plants, but with the addition of phosphite, transgenic tobacco plants easily outcompeted weeds (pictured).

Nature Biotechnol. <http://dx.doi.org/10.1038/nbt.2346> (2012)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

POLICY

Missile defence

The US missile-defence programme should end costly efforts to catch missiles in the early stages of launch and strengthen its ability to intercept them in mid-course, says *Making Sense of Ballistic Missile Defence*, a study from the National Academy of Sciences published on 11 September. Research into expensive space-based radars should be slashed, and efforts to distinguish warheads from dummies intensified, it says.

Climate-model plan

US climate modellers should work towards better ways of developing, interpreting and comparing climate and weather models, the National Research Council concludes in a report released on 7 September. A *National Strategy for Advancing Climate Modeling* recommends that the community try to standardize software platforms so that they are ready for the next generation of supercomputers. See go.nature.com/b2ymhi for more.

Nuclear future

The United Arab Emirates has started to build its first nuclear reactor, making it

JOURNALISM AWARD

Gayathri Vaidyanathan, who worked at *Nature* during a fellowship from the International Development Research Centre in Ottawa, Canada, has won the Evert Clark/Seth Payne Award for young journalists. The award cited 'The wheat stalker' (*Nature* **474**, 563–565; 2011) and 'The cultured chimpanzees' (*Nature* **476**, 266–269; 2011), and two pieces published elsewhere.



CHINA/GETTY IMAGES

Yellow river turns red

A stretch of the Yangtze River near Chongqing in southwest China turned red on 6 September. What caused the sudden change in the river's colour is still unclear, but freshwater scientists contacted by *Nature* say that it is unlikely to

be attributable to an algal bloom because such algae do not grow in oxygen-rich, fast-flowing river waters. Last December, the Jian River in northern China turned red as a result of dye pollution.

the first country to embark on a commercial nuclear programme since China in 1985. On 30 August, reactor unit 1 at Barakah was reported by the International Atomic Energy Agency in Vienna as being officially "under construction". The reactor is the first of four that will be supplied by a South Korean consortium at a total cost of US\$20 billion. The 1,400-megawatt pressurized water reactor is scheduled to begin operations in 2017.

Shale-gas risks

Drilling for shale gas might contaminate water resources, damage biodiversity and pollute the air — and such environmental risks are not adequately mitigated by existing legislation, says a study published by the European Commission on 7 September. A second study

notes that European shale gas would generate more greenhouse-gas emissions than conventional gas — but fewer than would be produced by importing gas into the European Union.

BUSINESS

Oil exploration

On 9 September, Shell began oil exploration in the Chukchi Sea off Alaska, the first offshore drilling in the sea for more than 20 years. The drilling is designed to reach an oil reservoir at a depth of 2,400 metres, but is unlikely to be finished this year: Shell's current permit is valid only until 24 September because of the threat of encroaching sea ice. Shell has requested an extension, arguing that sea ice is likely to arrive later this year. See go.nature.com/kjdtih for more.

RESEARCH

New species online

Animal species can now be described and named in electronic-only publications, the International Commission on Zoological Nomenclature announced on 4 September. Previously, descriptions of new animal species had to be published in print to be considered bona fide. Botanists endorsed electronic publication last year (see go.nature.com/ufq7sv). See page 178 and go.nature.com/pta96e for more.

Drilling depth

The deep-sea drilling vessel *Chikyu* has recovered rock samples from more than 2.1 kilometres beneath the sea floor, the Japan Agency for Marine-Earth Science and Technology announced on 6 September. *Chikyu* is drilling

off the coast of Japan north of the fault zone responsible for the 2011 Tohoku earthquake. Although oil wells frequently reach deeper, this is the first scientific expedition to retrieve core samples from such depths. See go.nature.com/vxrcrg for more.

Dengue trial

A dengue vaccine protected Thai schoolchildren from some forms of the mosquito-borne virus, reports a study published on 10 September (A. Sabchareon *et al.* *Lancet* <http://doi.org/jbq;2012>). Overall, 2.8% of 2,669 children who received the vaccine developed dengue, compared with 4.4% of the 1,333 in a control group, a result that was not statistically significant. But a secondary analysis found that the vaccine was 60–90% effective against three out of four genetic forms of dengue.

PEOPLE

Telescope chief

The world's largest radio-telescope project has chosen a director-general. On 5 September, the board of the Square Kilometre Array appointed Phil Diamond to oversee construction of the €1.5-billion (US\$1.9-billion) project, which will have sites in Australia and South Africa (see *Nature* **485**, 555–556;



2012). Diamond (pictured) is currently serving as the head of astronomy and space science at Australia's Commonwealth Scientific and Industrial Research Organisation in Marsfield. He replaces Michiel van Haarlem, who has served as interim director-general since December 2011.

Hauser disciplined

Marc Hauser, a former psychologist at Harvard University in Cambridge, Massachusetts, engaged in research misconduct, the US Office of Research Integrity concluded on 5 September. He neither admits nor denies the charges. See page 189 for more.

FUNDING

Football donation

The US National Football League (NFL) is donating US\$30 million to the Foundation for the National

Institutes of Health to support research into brain injuries. The money will help to establish a Sports and Health Research Program at the National Institutes of Health, to fund research on chronic traumatic encephalopathy and other sport-related brain injuries. Meanwhile, a study in *Neurology* (E. J. Lehman *et al.* *Neurology* <http://doi.org/h97;2012>) has reported that professional American football players are more likely than the general population to have Alzheimer's disease or the neurodegenerative disease amyotrophic lateral sclerosis as an underlying cause of death.

AWARDS

Balzan prizes

Kurt Lambeck, a geophysicist at the Australian National University in Canberra, has been awarded one of this year's two 750,000-Swiss-franc (US\$790,000) Balzan prizes in science for his research on sea-level changes and glacial cycles. Plant scientist David Baulcombe of the University of Cambridge, UK, won the other for his work on epigenetics.

Lasker award

The US\$250,000 Albert Lasker Basic Medical Research Award has been awarded to biologist Michael Sheetz of Columbia University in New York city,

COMING UP

14 SEPTEMBER

More details on the workings of the US National Institutes of Health's youngest centre, the National Center for Advancing Translational Sciences, are revealed at two key meetings in Bethesda, Maryland. The centre's advisory council and its Cures Acceleration Network Review Board both convene.

go.nature.com/zablws

17–21 SEPTEMBER

Nuclear safety and security are discussed at the annual general conference of the International Atomic Energy Agency in Vienna.

go.nature.com/dfaznk

biochemist James Spudich of the Stanford University School of Medicine in California and cell biologist Ronald Vale of the University of California, San Francisco, for their work on cytoskeletal motor proteins, which underlie cellular transport and muscle contraction. Winners of the award often go on to receive a Nobel prize.

EVENTS

Dawn departure

NASA's mission to explore the asteroid belt's two most massive bodies departed from its first target, Vesta, on 5 September. The Dawn probe had been mapping Vesta since July 2011. It is now heading for an early-2015 appointment with Ceres, the dwarf planet that is thought to harbour substantial amounts of subterranean ice. See go.nature.com/pgoftk for more.

► **NATURE.COM**

For daily news updates see:

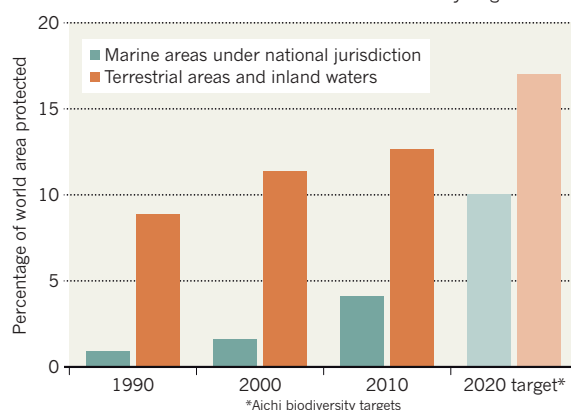
www.nature.com/news

TREND WATCH

The world's network of protected areas is growing rapidly, but not fast enough to meet the 2020 Aichi biodiversity targets, according to the trend detailed in the United Nations' *Protected Planet* report, released on 7 September at the IUCN World Conservation Congress in Jeju, South Korea. The proportion of area managed exclusively by governments has fallen, as conservation schemes based on local community management (often by indigenous peoples) play a growing part. See go.nature.com/dgve5x for more.

PROTECTED AREAS GROW — SLOWLY

The extent of the world's protected areas has increased 48% since 1990 — but more are needed to reach 2020 biodiversity targets.



NEWS IN FOCUS

PUBLIC HEALTH Europe seeks early warning of invading mosquitoes **p.187**

MISCONDUCT Guilty verdict settles Marc Hauser case — but not the questions **p.189**

ASTRONOMY Distorted galaxies scrutinized for dark energy **p.190**



OCEANOGRAPHY Alvin upgraded while projects founder **p.194**



E. BACCAGA/AGE FOTOSTOCK/ROBERT HARDING PICTURE LIBRARY

Larger areas of open ocean and warming surface waters will alter evaporation, cloudiness, precipitation and storm patterns across the Arctic and beyond.

CLIMATE

Ice loss shifts Arctic cycles

Record shrinkage confounds models and portends atmospheric and ecological change.

BY QUIRIN SCHIERMEIER

Before indifferent satellite eyes, the top of the world is undergoing a transformation. The Arctic ice pack, a primary indicator of climate change, has shrunk in recent weeks to an extent that no computer model and few scientists had thought possible.

After five years that all saw less ice than previously documented in the 34-year satellite record, this year's record loss has scientists questioning their models. They are also striving to understand the complex cascade of effects — from shifting weather patterns to displaced marine species — that the accelerating retreat could trigger.

The US National Snow and Ice Data Center

(NSIDC) in Boulder, Colorado, announced the record decline on 26 August, saying that the ice extent had dropped to 4.10 million square kilometres (see 'Going, going...'). The figure is 70,000 square kilometres less than the previous record low, set in 2007, and it came at least two weeks before the annual low is typically reached. According to the NSIDC, by 9 September that figure had dropped by another 14%, to around 3.52 million square kilometres.

The massive melt has occurred in relatively normal weather conditions, with only one strong summer storm to hasten the break-up of the pack ice. Mark Serreze, director of the

NSIDC, says that much of the Arctic pack is now thin first-year ice — frozen only last winter — which requires much less energy to break apart and disperse than multi-year ice. "We have entered a new regime," he says. "The sea ice is in such poor health in spring that large parts of it can't survive the summer melt season, even without boosts from extreme weather."

Computer models that simulate how the ice will respond to a warming climate project that the Arctic will be seasonally 'ice free' (definitions of this vary) some time between 2040 and the end of the century. But the observed downward trend in sea-ice cover suggests that summer sea ice could disappear completely as early as 2030, something that none of ►

NATURE.COM
For more on the
receding Arctic
ice, see:
go.nature.com/s5w2pr

► the models used for the next report by the Intergovernmental Panel on Climate Change comes close to forecasting¹.

"There's a tremendous spread between observations and model projections," says Serreze. "It might be that natural variability is larger than assumed, or perhaps models don't get the change in ice thickness right." Uncertainty also remains over the strength of various natural 'feedbacks'. For example, an exposed ocean is darker than an ice-covered surface and so absorbs more solar heat, causing yet more warming and melting.

A lack of fine detail about circulation patterns in the Arctic Ocean could also be throwing off the models. For example, a survey carried out in 2008 revealed 20 formerly unobserved eddies, each some 15 to 20 kilometres in diameter, in waters north of Canada. "Whether these are new features, and what role they might play for ocean-mixing processes, we don't know yet," says Yves Gratton, an oceanographer and Arctic researcher at the National Institute of Scientific Research in Montreal, Canada.

Ice loss could also accelerate if the ice pack's underlying waters warm up. Unlike in most of the world's oceans, the coldest water in the Arctic, at -1°C to -2°C , is at the surface; below a depth of 200–300 metres, saltier and warmer water of about 1°C flows in from the Atlantic. The cold surface layer — called the halocline — isolates the sea ice from the warmer water below.

But the halocline is vulnerable to warming from above, says Henning Bauch, a marine geologist at the GEOMAR research centre in Kiel, Germany. A thinning halocline — something that has not yet been observed — would not only jeopardize the sea ice but could also melt the carbon-rich permafrost beneath shallow coastal waters², releasing greenhouse gases into the atmosphere.

Meanwhile, Arctic biology is already changing, as the retreat and thinning of sea ice allows more sunlight to penetrate the upper ocean and deprives certain species of habitat, says Jorgen Berge, a marine biologist at the University of Tromsø in Norway. For example, the

GOING, GOING ...
Sea-ice coverage over the Arctic Ocean as measured on 26 August — a record-breaking low.



dominant Arctic zooplankton — the copepods *Calanus hyperboreus* and *C. glacialis* — are being replaced by Atlantic *C. finmarchicus*. Meanwhile, Arctic cod (*Arctogadus glacialis*) is increasingly being out-competed by its larger Atlantic cousin *Gadus morhua*³.

Other species are showing unexpected resilience. In January, a research cruise into the pack ice north of Svalbard, Norway, found evidence that *Apherusa glacialis* — amphipods, or small crustaceans, that feed on algae growing beneath the ice and are the main food source for many Arctic sea birds — are less dependent on extensive sea ice than previously thought. Just as some coastal organisms drift with the tide to retain their positions in river estuaries, these amphipods can sink to depths at which pole-ward ocean currents sweep them farther north, where they can find and colonize surviving sea ice⁴.

"This looks like an adaptive trait that allows key Arctic-ice fauna to persist in changing sea-ice environments," says Berge, who led the study. He points out that the Arctic Ocean has been ice-free during the summer several times during the past 2.5 million years, and that as recently as 8,000–6,500 years ago, during a period known as the Holocene Thermal

Maximum, much more of the Arctic was ice-free than it is today⁵ — with no mass-extinction events known to have occurred.

Farther afield, this year's record sea-ice melt might foreshadow a harsh winter in parts of Europe and North America. Recent research, although preliminary, suggests a connection between late-summer Arctic sea-ice extent and the location of areas of high and low atmospheric pressure over the northern Atlantic. The highs and lows can remain relatively fixed for weeks, shaping storm tracks and seasonal weather patterns such as extended cold surges⁶.

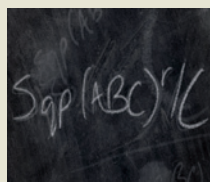
Ralf Jaier, a climate scientist at the Alfred Wegener Institute for Polar and Marine Research in Potsdam, Germany, found a significant correlation in 1989–2011 meteorological data between late-summer Arctic sea-ice extent and atmospheric-pressure anomalies that favour extreme weather such as prolonged cold snaps in winter⁷. He reasons that in autumn, the open Arctic Ocean sheds heat to the high-latitude atmosphere. The warming tends to reduce the large-scale atmospheric-pressure gradient and weakens the dominant westerly winds in the Northern Hemisphere. Those winds normally sweep warm, moist Atlantic air to western Europe; their weakening leaves the region more prone to persistent cold.

"The impacts will become more apparent in autumn, once the freeze-up is under way and we see how circulation patterns have influenced the geographical distribution of sea ice," says Judith Curry, a climate researcher at the Georgia Institute of Technology in Atlanta. But, she adds, "We can probably expect somewhere in the mid/high latitudes of the Northern Hemisphere to have a snowy and cold winter." ■

1. Stroeve, J. C. *et al.* *Geophys. Res. Lett.* **39**, L16502 (2012).
2. Vonk, J. E. *et al.* *Nature* **489**, 137–140 (2012).
3. Schiermeier, Q. *Nature* **446**, 133–135 (2007).
4. Berge, J. *et al.* *Biol. Lett.* <http://dx.doi.org/10.1098/rsbl.2012.0517> (2012).
5. Funder, S. *et al.* *Science* **333**, 747–750 (2011).
6. Liu, J., Curry, J. A., Wang, H., Song, M. & Horton, R. M. *Proc. Natl Acad. Sci. USA* **109**, 4074–4079 (2012).
7. Jaier, R., Dethloff, K., Handorf, D., Rinke, A. & Cohen, J. *Tellus A* **64**, 11595 (2012).

MORE ONLINE

TOP STORY

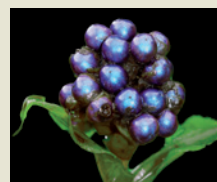


'Astounding' solution to complex maths puzzle
go.nature.com/2lz9lk

MORE NEWS

- Science wins over creationism in South Korea go.nature.com/abfuyd
- Rabbit grimaces unlock animal pain go.nature.com/qenfku
- Dietary supplements may offer a way to treat rare form of autism go.nature.com/gbevdf

SLIDESHOW



Glittering African fruit has nature's brightest hue
go.nature.com/o39pqy



The Asian tiger mosquito (*Aedes albopictus*) can transmit at least 22 different viruses.

DISEASE CONTROL

Europe on alert for flying invaders

Spread of disease-carrying mosquitoes prompts guidelines for boosting surveillance.

BY DECLAN BUTLER

The buzz of tiny wings has been growing ever louder in Europe over the past decade, as invasive mosquitoes carrying exotic diseases such as chikungunya and dengue fever have taken up residence around the continent. Now Europe's entomologists and public-health experts are joining forces to defend the region against the growing threat.

On 28 August, the European Centre for Disease Prevention and Control (ECDC) in Stockholm released guidelines that should help European countries to develop mosquito surveillance, which had become mostly unnecessary after malaria was largely eradicated from the continent in the mid-twentieth century. But the need for better vigilance has been growing again since the late 1990s, as the increased globalization of trade gives species such as the Asian tiger mosquito (*Aedes albopictus*) more opportunities to penetrate into Europe.

The surveillance data will feed into the three-year-old VBORNET consortium, a

Europe-wide network of some 400 medical entomologists, microbiologists and public-health experts that is funded by the ECDC. The wealth of data they are assembling should help researchers to better understand how invasive mosquitoes are spreading and the risks they pose, and so to develop control strategies.

The latest guidelines provide an A–Z of mosquito surveillance, covering topics ranging from what sort of mosquito traps should be used in particular circumstances to resource management and policy-making advice. Francis Schaffner, head of integrated pest management at the consultancy Avia-GIS in Zoersel, Belgium, and one of the lead authors of the guidelines, hopes that they will encourage scientists, local authorities and governments to carry out more comprehensive surveillance. By recommending standardized

➔ **NATURE.COM**
Read more on
vector control in this
Outlook:
go.nature.com/udwjuu

surveillance reporting, the guidelines should also allow data generated by different researchers in different countries to be seamlessly combined

into pan-European databases, he says.

A key target of surveillance is the Asian tiger mosquito, which can transmit at least 22 different viruses, including chikungunya and dengue. After it was reported in Albania in 1979, this native of southeast Asia spread to Italy in 1990 and has now expanded along the Mediterranean coast to nations such as France and Spain, as well as around the Black Sea (see 'Multiplying mosquitoes') (J. M. Medlock *et al.* *Vector-Borne Zoonotic Dis.* **12**, 435–447; 2012). At least five other invasive mosquitoes capable of spreading diseases are also gaining ground in Europe, including *Aedes aegypti*, which can transmit dengue and other diseases.

The international tyre business brings most of the unwanted arrivals. Mosquitoes can breed in small volumes of water, such as rainwater in tyres, and invasive forms have drought-resistant eggs that enable stowaways to survive long-distance transport before being revived by rain once they reach their destinations. The invaders can also hitch a ride on imported plants, such as lucky bamboo (*Dracaena* species). In the future, climate change could accelerate the spread of these subtropical and tropical species: predictive models suggest that invasive mosquitoes could find suitable habitats in northern Europe by 2030.

In 2007, Italy saw an outbreak of several hundred cases of chikungunya, which causes nausea, vomiting and debilitating joint pain. Dengue fever, which causes similar symptoms and can be fatal, hit a handful of people in France and Croatia in 2010. And there has been a steady rise in cases of West Nile virus, the fever that is currently notching up a record year for infections in the United States.

Surveillance remains patchy or non-existent in many of the most affected countries around Europe, including Portugal, Greece, Turkey and those bordering the Black Sea. Monitoring insects is expensive and time-consuming, and governments that are preoccupied with existing health challenges may be reluctant to put funds towards tracking potential future threats. But one underlying message of the report is that early surveillance is ultimately cheaper than dealing with an established pest, which can require widespread spraying with insecticides.

The latest guidelines provide an A–Z of mosquito surveillance.

In the Netherlands, for example, Asian tiger mosquitoes are introduced every year, but surveillance and control measures have so far prevented them from becoming established, says Marieta Braks, an entomologist at the National Institute for Public Health and the Environment in Bilthoven, and leader of VBORNET's strategic consultation group. The Dutch authorities pay close attention to companies that import tyres from countries harbouring dangerous mosquitoes, and that stock tyres outdoors, ►

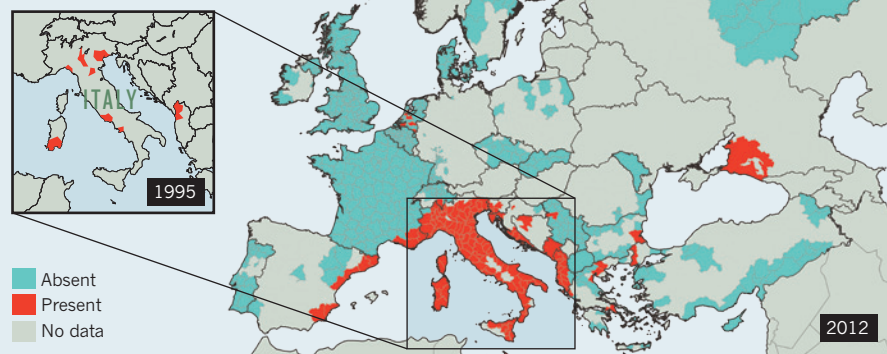
► where rain could allow mosquito eggs to hatch. In countries such as Italy, where the tiger mosquito is already widespread, surveillance is instead geared more towards keeping numbers under control.

VBORNET's geographical information system allows scientists to combine surveillance data with environmental factors, including temperature and humidity, to better track the main invasive mosquito species. Models based on these combined data can reveal, for example, how environmental conditions can affect mosquitoes' ability to transmit pathogens, and so provide information on the likelihood of outbreaks in various regions.

The maps may also perform a useful peer-pressure function, says Braks, by highlighting where researchers and authorities have failed to supply and share data. "What we hope is that no one will want to be a blank on the map." ■

MULTIPLYING MOSQUITOES

Data from June 2012 show that the Asian tiger mosquito (*Aedes albopictus*) has spread rapidly across southern Europe over the past decade.



FUNDING

EU battles over research billions

Politicians will spend the next few months negotiating cuts to Horizon 2020's budget.

BY NATASHA GILBERT

Tens of billions of euros are at stake as negotiations ramp up to shape Europe's next seven-year research programme. The discussions will cover familiar divisions over applied versus basic research and conflicting national agendas, but the continent's ongoing financial problems will add an extra measure of anxiety.

The European Commission proposed the general size and shape of the Horizon 2020 research and innovation programme in November 2011 (see *Nature* <http://doi.org/dtvtbx>; 2011). European Union (EU) member states and members of the European Parliament will negotiate and agree on the budget and final

details of the programme, which will run from 2014 to 2020, over the next 15 months.

Last year, the commission requested €77.6 billion (US\$99 billion) for Horizon 2020, out of a total European budget for 2014–20 of about €1 trillion (at constant 2011 prices). But many members of the European Parliament (MEPs) want research to get €100 billion, twice as much as was allocated to Horizon 2020's predecessor, the Seventh Framework Programme (FP7), which runs from 2007 until 2013 (see 'Budget breakdown'). Most scientists and EU science officials say that such a large jump is unrealistic, and there are growing fears that even the lower budget proposed by the commission is unlikely, given Europe's shaky economy.

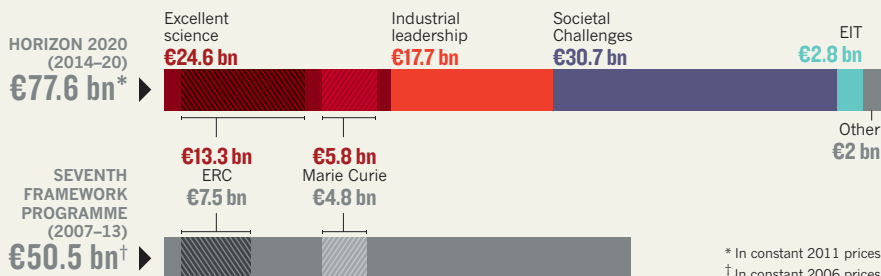
At a meeting on 30 August, high-level ministers said that cuts to the proposed EU budget are "inevitable", implying that spending on Horizon 2020 will be eroded. Ministers are aiming to reach a consensus on the size and shape of the EU budget by the end of the year, but in the meantime, MEPs, scientists and member-state officials will fight over which of Horizon 2020's ambitions should bear the brunt of any budget reductions. The first sortie will come on 17–18 September, when MEPs on the European Parliament's Industry, Research and Energy (ITRE) committee will begin discussing around 3,000 suggested amendments to the Horizon 2020 budget.

Most likely to feel the pinch is the European Institute of Innovation and Technology (EIT), a virtual institute headquartered in Budapest and made up of scientific 'communities' that conduct research on themes such as climate change and sustainable energy. The institute also runs postgraduate education programmes and knowledge-transfer activities to disseminate the fruits of research to businesses and policy-makers (see *Nature* 440, 8–9; 2006). The EIT concept was developed and championed by the European Commission, which now suggests that the institute's budget should win €2.8 billion from Horizon 2020, 3.6% of the total research budget and vastly more than the €309 million in European funds allocated to it between 2008 and 2013.

However, few outside the commission want the EIT to claim so much; some argue that the

BUDGET BREAKDOWN

The proposed budget for Europe's Horizon 2020 research-funding programme, the successor to its Seventh Framework Programme, would boost the European Research Council (ERC), the Marie Curie Actions career-development programme and the European Institute of Innovation and Technology (EIT).



institute has yet to prove its scientific worth. Teresa Madurell, a Spanish social-democrat MEP on the ITRE committee, suggests that the EIT should get no more than 3.1% of Horizon 2020's funds. Paul Rübiger, an Austrian Christian Democrat MEP who also sits on the committee, wants to trim that even further, suggesting that the EIT's remit be reduced to knowledge-transfer activities only.

The EIT's critics say that the money would be better spent on initiatives with a proven track record. The most popular programmes include the European Research Council (ERC), which dispenses excellence-based grants for frontier science, and the Marie Curie Actions programme, which provides career-development grants to young researchers. The commission proposed that the ERC's budget should rise by 77% to €13.3 billion, but Luke Georgioui, who studies European research policy at the University of Manchester, UK, expects opposition from member states in eastern Europe, which could hold up the budget negotiations. Scientists in those countries often lose out to those from research powerhouses such as the United Kingdom and Germany when ERC grants are allocated. Some say that the ERC should also fund those scientists with the 'potential for excellence'. Vicky Ford, a British Conservative member of the ITRE committee, counters that ERC funding should go only to excellent research.

The ERC tends to fund individual researchers, so during FP7 larger collaborations have relied on separate funding streams. But some scientists are concerned that funding for applied research has squeezed out support for basic science in these programmes. Over the past four years, calls for collaborative proposals in areas of basic science such as epigenetics and protein regulation have become less frequent, says Karin Metzloff, executive director of the European Plant Science Organisation in Brussels. She hopes that the ITRE committee meeting can help to make basic-science projects a bigger priority in Horizon 2020's collaborative-research programme, dubbed Societal Challenges.

Overall, research is likely to be spared the EU budget's most severe cuts, because politicians recognize that investing in science is central to boosting economic growth. Indeed, countries such as Spain, which has slashed domestic science spending owing to the financial crisis, will become more reliant on Horizon 2020 funds. That may even prompt them to give ground in negotiations on contentious issues such as human embryonic stem-cell research, says Georgioui. "They need to replace their national funding shortage," he says. ■



Marc Hauser left Harvard University in 2011 after the institution found him guilty of misconduct.

RESEARCH ETHICS

Misconduct ruling is silent on intent

Psychologist Marc Hauser admits errors but not fraud.

BY EUGENIE SAMUEL REICH

Mavens of research ethics often insist that there is a clear difference between sloppy science and scientific fraud. But if ever there was a case that blurs that line, it is that of Marc Hauser, a high-flying evolutionary psychologist who resigned from Harvard University in Cambridge, Massachusetts, in 2011, after the university found him guilty of misconduct.

In a finding published on 6 September in the *Federal Register*, the US Office of Research Integrity (ORI) in Rockville, Maryland, which oversees science funded by the National Institutes of Health (NIH), has also concluded that Hauser "engaged in research misconduct". The office found that he: "published fabricated data" in the journal *Cognition* in 2002¹; "falsified" results in an unpublished study; and made false statements in four other instances. But the ORI did not state whether Hauser acted intentionally — its definition of misconduct also covers cases in which misrepresentations are "reckless".

That has provided enough wiggle room for both supporters and critics to claim the finding as a vindication. "It isn't a strong case for intentional misconduct. It seems it's peccadilloes," says Pierre Pica, a linguist at the French National

Center for Scientific Research in Paris, who has criticized the investigation. But Gerry Altmann, a psychologist at the University of York, UK, and editor-in-chief of *Cognition*, disagrees. "The number of instances [in the investigated body of work] suggests it is extremely unlikely it was not intentional," he says.

Harvard had found Hauser guilty of eight counts of misconduct in 2010, after a three-year internal investigation of some of his studies of cognition in non-human primates. To the annoyance of many scientists, the university did not release its findings, so exactly what Hauser had done remained unclear. The ORI provides details on six counts of misconduct, three of which were published¹⁻³.

In a statement sent to *Nature*, Hauser acknowledges mistakes and errors, but not scientific fraud or research misconduct, which the ORI says he neither admits nor denies. "Although I have fundamental differences with some of the findings in the ORI report," Hauser says, "I acknowledge that I made mistakes. I let important details get away from my control, and as head of the lab, I take responsibility for all errors made within the lab, whether or not I was directly involved." No one else who worked in Hauser's lab has been charged with misconduct.

The ORI often resolves allegations ►

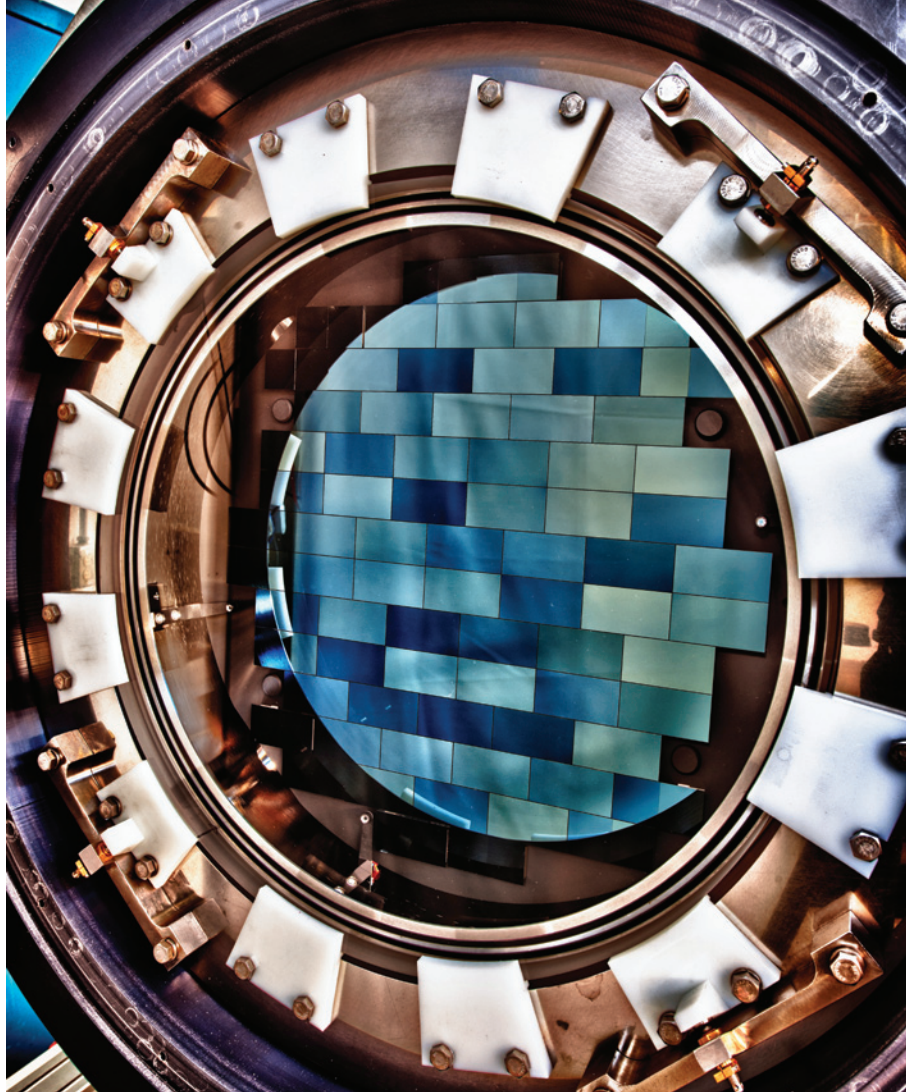
► against scientists through a settlement agreement, which helps the reputedly overstretched office to save on legal resources. Hauser's settlement means that he escapes the harshest possible punishment — a ban on future NIH funding — and must instead submit a supervisory plan that will ensure integrity if he applies to the agency for any future research grants. (The agency spent at least US\$790,000 on grants that funded work affected by the misconduct.)

Bennett Galef, an expert in animal behaviour at McMaster University in Hamilton, Canada, reviewed evidence used in the Harvard investigation at the request of Hauser's lawyer, and questions the ORI's use of the word "fabricated" to describe the errors in the *Cognition* paper¹. "That is a conceivable read of what happened, but it's slanted," he says. According to Galef, lab records show that in the study, which examined habituation to sound patterns, two groups of tamarin monkeys were played the same stimuli instead of different ones, as the paper reported. Galef says that this was because of an error in the computer program that played the stimuli, which the authors plausibly didn't know about because the protocol called for them to be blinded to the stimuli at the time they coded the monkeys' responses. "It was a big mistake. It was definitely a disaster, but whether there was any fabrication, you don't know," says Galef.

ORI officials familiar with Hauser's case could not comment because of government privacy rules. The inspector-general of the National Science Foundation, which also funded Hauser, is still looking into one of the misconduct counts found by Harvard; the remaining instance was referred to the ORI but has been deemed outside its statute of limitations. Harvard says that the graduate students and post-docs who were working in Hauser's lab have been relocated to other labs within the university.

Hauser has now begun a different chapter in his life, working with at-risk youth. His website states that he has co-founded a company called Gamience, which develops computer games to teach self-control and other cognitive skills, and is working with a non-profit company that serves schools on Cape Cod, called the Cape Cod Collaborative, in Bourne, Massachusetts. "This work is deeply satisfying and I look forward to making new contributions to human welfare, education and the role of scientific knowledge in understanding human nature," Hauser says. ■

1. Hauser, M. D., Weiss, D. & Marcus, G. *Cognition* **86**, B15–B22 (2002).
2. Hauser, M. D., Glynn, D. & Wood, J. *Proc. R. Soc. B* **274**, 1913–1918 (2007).
3. Wood, J. N., Glynn, D. D., Phillips, B. C. & Hauser, M. D. *Science* **317**, 1402–1405 (2007).



R. HAHN/FERILAB

The Dark Energy Survey camera will investigate millions of galaxies for the subtle effects of weak lensing.

COSMOLOGY

Cameras to focus on dark energy

A pair of detectors that measure minute distortions in images of distant galaxies will probe the riddle of cosmic acceleration.

BY ERIC HAND

Even the best pictures of a distant galaxy are a bit lopsided. But this is an attribute, not a bug. Because mass distorts space-time, light coming from distant galaxies is bent as it passes through intervening shoals of invisible matter, leaving the images of these distant objects minutely sheared and stretched.

Two astronomical surveys now scheduled to come online seek to take advantage of this effect, which is known as weak gravitational lensing. The surveys aim to use the technique to get a firmer handle on dark energy, the mysterious force that is apparently speeding up the

expansion of the Universe. By observing the patterns of distortions across large swathes of sky (see 'Falling into line'), astronomers hope to map the density and distribution of dark matter, the web-like invisible scaffolding around which visible matter is thought to have first coalesced. Then, by looking at changes in this hidden web across cosmic time, they hope to discern the imprint of dark energy.

➔ **NATURE.COM**
Read more about
the search for
dark energy at:
go.nature.com/cz3f64

Observers already study the effects of dark energy by tracking cosmic landmarks: the standard candles of distant supernovae and the

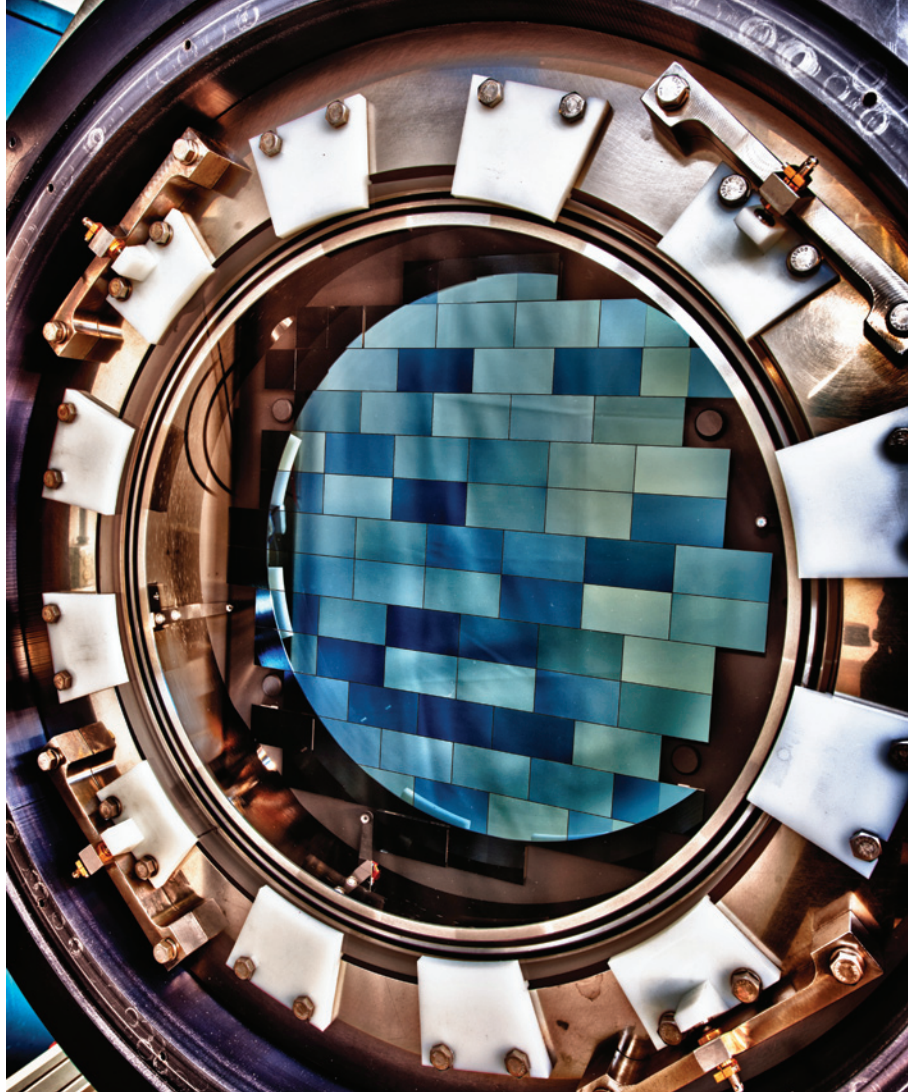
► against scientists through a settlement agreement, which helps the reputedly overstretched office to save on legal resources. Hauser's settlement means that he escapes the harshest possible punishment — a ban on future NIH funding — and must instead submit a supervisory plan that will ensure integrity if he applies to the agency for any future research grants. (The agency spent at least US\$790,000 on grants that funded work affected by the misconduct.)

Bennett Galef, an expert in animal behaviour at McMaster University in Hamilton, Canada, reviewed evidence used in the Harvard investigation at the request of Hauser's lawyer, and questions the ORI's use of the word "fabricated" to describe the errors in the *Cognition* paper¹. "That is a conceivable read of what happened, but it's slanted," he says. According to Galef, lab records show that in the study, which examined habituation to sound patterns, two groups of tamarin monkeys were played the same stimuli instead of different ones, as the paper reported. Galef says that this was because of an error in the computer program that played the stimuli, which the authors plausibly didn't know about because the protocol called for them to be blinded to the stimuli at the time they coded the monkeys' responses. "It was a big mistake. It was definitely a disaster, but whether there was any fabrication, you don't know," says Galef.

ORI officials familiar with Hauser's case could not comment because of government privacy rules. The inspector-general of the National Science Foundation, which also funded Hauser, is still looking into one of the misconduct counts found by Harvard; the remaining instance was referred to the ORI but has been deemed outside its statute of limitations. Harvard says that the graduate students and post-docs who were working in Hauser's lab have been relocated to other labs within the university.

Hauser has now begun a different chapter in his life, working with at-risk youth. His website states that he has co-founded a company called Gamience, which develops computer games to teach self-control and other cognitive skills, and is working with a non-profit company that serves schools on Cape Cod, called the Cape Cod Collaborative, in Bourne, Massachusetts. "This work is deeply satisfying and I look forward to making new contributions to human welfare, education and the role of scientific knowledge in understanding human nature," Hauser says. ■

1. Hauser, M. D., Weiss, D. & Marcus, G. *Cognition* **86**, B15–B22 (2002).
2. Hauser, M. D., Glynn, D. & Wood, J. *Proc. R. Soc. B* **274**, 1913–1918 (2007).
3. Wood, J. N., Glynn, D. D., Phillips, B. C. & Hauser, M. D. *Science* **317**, 1402–1405 (2007).



R. HAHN/FERMI LAB

The Dark Energy Survey camera will investigate millions of galaxies for the subtle effects of weak lensing.

COSMOLOGY

Cameras to focus on dark energy

A pair of detectors that measure minute distortions in images of distant galaxies will probe the riddle of cosmic acceleration.

BY ERIC HAND

Even the best pictures of a distant galaxy are a bit lopsided. But this is an attribute, not a bug. Because mass distorts space-time, light coming from distant galaxies is bent as it passes through intervening shoals of invisible matter, leaving the images of these distant objects minutely sheared and stretched.

Two astronomical surveys now scheduled to come online seek to take advantage of this effect, which is known as weak gravitational lensing. The surveys aim to use the technique to get a firmer handle on dark energy, the mysterious force that is apparently speeding up the

expansion of the Universe. By observing the patterns of distortions across large swathes of sky (see 'Falling into line'), astronomers hope to map the density and distribution of dark matter, the web-like invisible scaffolding around which visible matter is thought to have first coalesced. Then, by looking at changes in this hidden web across cosmic time, they hope to discern the imprint of dark energy.

➔ **NATURE.COM**
Read more about
the search for
dark energy at:
go.nature.com/cz3f64

Observers already study the effects of dark energy by tracking cosmic landmarks: the standard candles of distant supernovae and the

standard rulers of ripples in the distribution of galaxies. But those techniques reveal only how the Universe's expansion rate has changed because of dark energy. Weak lensing offers an extra prize: the distortions hold clues to the action of gravity at far-off locales. The technique can thus address an increasingly pressing question for theorists: what if the Universe's accelerated expansion is not the result of dark energy's mysterious anti-gravitational force, but is instead a sign that gravity behaves differently in different parts of the Universe?

"It provides unfettered insight into the properties of gravity that we can't get by other techniques," says Rachel Bean, an astrophysicist at Cornell University in Ithaca, New York. "Its potential is massive, but so are the challenges."

Unlike strong gravitational lensing, in which a major concentration of mass markedly distorts the image of a single galaxy, the subtle effects of weak lensing can be detected only in detailed surveys of millions of galaxies. One of the new efforts will use the Hyper Suprime-Cam (HSC), which achieved first light on 28 August on the 8.2-metre Subaru telescope in Hawaii. By 2018, it aims to have imaged 10 million galaxies over a 1,500-square-degree swathe of the sky, says Satoshi Miyazaki of the National Astronomical Observatory of Japan in Mitaka, who is the principal investigator of the HSC survey.

A rival project, the Dark Energy Survey (DES), is set to start operating later this month on the 4-metre Blanco telescope in Chile, says its principal investigator, Josh Frieman of the Fermi National Accelerator Laboratory in Batavia, Illinois. Being the main project for the Blanco telescope, the DES will get more observing nights than the HSC, allowing it to image 300 million galaxies over 5,000 square degrees by 2018. But its smaller telescope means that it will not be able to peer as deeply into the distant Universe as the HSC survey.

Even with these massive digital cameras — 870 million pixels for the HSC and 570 million pixels for the DES — weak lensing "is a very, very difficult measurement to make," says Bean. To detect the subtle distortions caused

by lensing, for example, astronomers must subtract aberrations caused by the optics of the telescope and by Earth's atmosphere.

Doing the measurements from space could help. The European Space Agency is planning to launch a dark-energy probe called Euclid in 2019. And in the United States, an as-yet-unfunded dark-energy mission called the Wide-Field Infrared Survey Telescope, which would dedicate much of its time to weak-lensing measurements, was ranked as the top priority for space-based astronomy in a 2010 decadal survey.

Some astronomers doubt that the technique can outperform supernova surveys and measurements of baryonic acoustic oscillations — ripples in cosmic structure dating from several hundred thousand years after the Big Bang. "I think the jury's still out," says David Schlegel,

principal investigator of the Baryon Oscillation Spectroscopic Survey at the Lawrence Berkeley National Laboratory in Berkeley, California. He points out, for

"The technique's potential is massive, but so are the challenges."

instance, that a key assumption in weak lensing — that galaxies are oriented randomly — is turning out not to be true, forcing astronomers to correct for a systematic effect.

But others say that it is important to take diverse approaches to the dark-energy mystery. Chris Stubbs, an astrophysicist at Harvard University in Cambridge, Massachusetts, says that, a decade ago, astronomers were intent on simply proving that the effect was real. Now there is little doubt, and they need techniques such as weak lensing to explore more-detailed questions — for instance, whether the effect varies with distance, direction or the density of matter, and whether the effects of dark energy can be distinguished from those of variable gravity. "We are only now entering an era where we're attacking the dark-energy problem with instruments and experiments that were explicitly optimized for that purpose," says Stubbs. ■

FALLING INTO LINE

When light from a randomly oriented group of galaxies passes through a region influenced by the gravity of a large concentration of dark matter, the galaxy images show a slight alignment along a particular direction. Such distortions can reveal the changing concentrations of matter at different cosmic times.





Soaring sales of antibiotics at Indian pharmacies are compounding drug-resistance problems.

PUBLIC HEALTH

India moves to tackle antibiotic resistance

Drug regulator aims to restrict over-the-counter sales.

BY ERICA WESTLY IN CHENNAI, INDIA

Kumar Prabhash's voice is sombre as he describes how he is losing a key weapon in the fight against hospital infections. At the Tata Memorial Centre in Mumbai, where the oncologist works, at least half of bacterial samples from patients with infections are resistant to carbapenems — a class of 'second-line' antibiotics used to treat infections that are already resistant to other drugs. Just a few years ago, the resistance rate in such samples was only 30%. "That is really scary for us," says Prabhash.

Antibiotic resistance is a problem worldwide, but is particularly worrying in India, where hospital standards are inconsistent and antibiotics are readily available over the counter at pharmacies. Prabhash's story was all too familiar to the physicians who heard it at India's first joint medical-society meeting on antibiotic resistance, held on 24 August in Chennai. The symposium is part of an ongoing campaign by Indian clinicians for a national policy to curb misuse of these vital drugs.

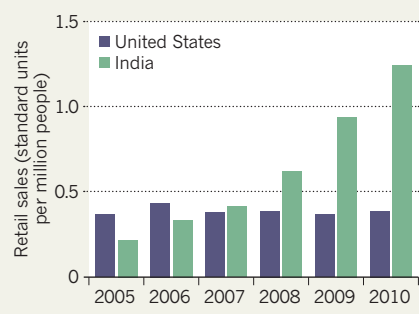
"As physicians, we have a responsibility for the future of India," said J. A. Jayalal, who represented the Indian Medical Association at the

meeting. "If we're not raising our voices about this, future generations will not forgive us."

Last week, physicians moved a step closer to their goal when India's drug regulators announced a plan that would put tight restrictions on the sale of antibiotics. Carbapenems and many other antibiotics are already on a list of 536 drugs in India that require a prescription. But studies have shown that such drugs are easy to purchase at retail pharmacies

ANTIBIOTICS FOR ALL

Carbapenem antibiotics are increasingly available without prescription in India, compared with Western nations. This is driving the emergence of microbial resistance to the drugs in the country.



without a physician's signature (U. P. Rathnaker *et al.* *J. Clin. Diagn. Res.* 6, 951–954 (2012) and A. Kotwani *et al.* *J. Clin. Pharm. Ther.* 37, 308–312; 2012). Under the new regulations, carbapenems and other antibiotics with declining efficacy would be treated as a special category of prescription drugs. They would carry prominent red labels, and surprise inspections by the regulator would check whether pharmacies were selling them over the counter.

Drug misuse is a key factor that leads to antibiotic resistance, says Arjun Srinivasan, a medical epidemiologist at the US Centers for Disease Control and Prevention in Atlanta, Georgia. A patient's risk of developing a life-threatening infection that is resistant to carbapenems is 15 times greater if he or she has been exposed to the drug class before, he notes. "Improved antibiotic use will be essential to combat drug resistance," he adds.

There is ample evidence of excessive carbapenem use in India. A survey (go.nature.com/rpsg2r) published earlier this year by the Center for Disease Dynamics, Economics and Policy, a non-profit health-research organization headquartered in Washington DC, found that Indian pharmacy sales of carbapenems increased by nearly six times between 2005 and 2010 (see 'Antibiotics for all').

Limited access to basic medical care is one reason for the overuse of carbapenems and other second-line antibiotics. About 70% of India's 1.2 billion citizens live in rural areas, where, despite government efforts, hospitals are often understaffed and lacking basic equipment and medication. Rather than relying on physicians, many rural patients turn to local pharmacists and whatever drugs they have in stock.

In 2010, India's government created a task force to develop a national antibiotic policy. But last year, Indian health minister Ghulam Nabi Azad rejected the task force's recommendations to ban over-the-counter sales of antibiotics and prohibit the use of carbapenems and other second-line antibiotics except in major hospitals. Azad argued that the bans could have a devastating effect on health in rural areas.

The health ministry did not respond to *Nature's* questions about the latest recommendation from India's drug regulator, but a health-ministry representative told the *Times of India* that the office did intend to approve the antibiotic restrictions, possibly as early as this week.

Infectious-disease specialist Abdul Ghafur, who organized the joint society meeting in Chennai, hopes to use momentum from the latest restrictions to push for a national antibiotic policy to regulate India's 24,000 hospitals and hundreds of thousands of physicians, as well as its pharmacies. "I am now trying to follow up with the health minister's office to arrange a meeting with the various medical societies," he says. "There are still many more steps ahead." ■

P. VISHWANATHAN/THE NEW YORK TIMES/REX/EVERETT

SOURCE: CODEP



DIVE MASTER

The US flagship submersible *Alvin* is getting a partial upgrade. But deep-sea exploration faces some rough water.

BY RICHARD MONASTERSKY

At an industrial hangar in Woods Hole, Massachusetts, a plush toy dog stands guard next to a 5-tonne titanium sphere. Polished to a high shine, the 2.2-metre orb reflects the lights overhead and the faces of passing scientists, who come to ogle the US\$11-million engineering marvel — a container strong enough to ferry researchers 6,500 metres down to the bottom of the ocean.

The sphere will be the new heart of the *Alvin* submersible, the United States' workhorse in the deep sea. Over its five-decade career, *Alvin* has dived to the sea floor more than any other research submersible and has made some of the most important oceanographic discoveries of the past few generations. But in the past few years, it has been showing its age,

and other countries have surged past with vehicles that can venture far deeper than *Alvin*'s current limit of 4,500 metres. In 2010, *Alvin*'s caretakers at the Woods Hole Oceanographic Institution (WHOI) took the venerable submersible out of commission for a \$40-million redesign.

The massive upgrade was a long time coming. "I started here in 1996 and there was talk about building a new sub then," says Bruce Strickrott, an *Alvin* pilot and expedition leader at WHOI, who is managing the sub's reconstruction.

This month, Strickrott and his crew will start to rebuild the sub around the passenger sphere, readying it for a series of final tests off the coast of Bermuda by March. Assuming it passes, *Alvin* will resume a full schedule of scientific dives, joining a suite of tools that are giving US researchers unprecedented ways to study the deep sea — from manned subs like *Alvin* to remotely controlled robots and autonomous mappers.

But after celebrating the revamped sub's maiden voyage, researchers will wake up to some sobering realities. Budget problems have constrained the current *Alvin* upgrade, so researchers must wait for the completion of a second construction phase, in 3–5 years, before they can take advantage of the greater depth limit allowed by the new sphere. That leaves the United States well behind France, Russia, Japan and now China, which sent its research submersible below 7,000 metres in June. And as wealthy entrepreneurs such as film director James Cameron capture public attention with dives in privately built subs, the United States is making painful cuts to many of its deep-sea assets. The administration of President Barack Obama is proposing to cut off funding next year for a pair of manned submersibles in Hawaii that can dive to 2,000 metres.

"America has one submersible capable of reaching Titanic-ish depths — in

The *Alvin* submersible was taken out of service for upgrades in 2010 and will dive again next year.

L WILLS/WOODS HOLE OCEANOGRAPHIC INSTITUTION

the couple-of-miles range — which these days isn't that deep," says Cameron. "What is the future of deep submergence research?"

That future once looked boundless. On 17 February 1977, *Alvin* skimmed along the ocean bottom about 400 kilometres northeast of the Galapagos Islands. Days before, sensors towed by a ship had identified a spike in water temperatures near the sea floor. *Alvin* explored the region and brought researchers face to face, for the first time, with hydrothermal vents, fissures in the sea floor that spew streams of warm, shimmering brine into the near-freezing water. In what was supposed to be a barren desert at the ocean bottom — far from the reaches of life-powering sunlight — *Alvin* revealed oases of clams, crabs, shrimp and alien-like white tube worms crowned with crimson tops.

Two years after the Galapagos dives, an *Alvin* expedition witnessed a more violent scene: jets of black fluids, superheated to 400 °C, shooting out of gigantic chimneys of rock, which were dubbed black smokers. Later, in the geologically quieter Gulf of Mexico, *Alvin* found vents that were releasing cooler methane-rich fluids. These 'cold seeps' also support communities of animals, including slow-growing tube worms thought to be hundreds of years old — some of the most ancient creatures alive on the planet. The discovery of such ecosystems, powered by microbes that harness chemical energy from vent waters, raised questions about where life originated on Earth and where it may reside on other worlds.

Alvin's string of accomplishments silenced early critics — some of the top geoscientists of the 1960s and 1970s — who initially regarded the submersible as little more than a toy. And ever since, the vehicle has been in high demand, trekking to the sea floor 100–150 times a year. Although some countries can dive more deeply, *Alvin* "is the most reliable submersible there is," says Susan Humphris, a geochemist at WHOI and principal investigator of the *Alvin* overhaul. "We need a workhorse at the bottom of the ocean. That's what *Alvin* has been and that's what *Alvin* will continue to be."

AT THE HELM

Cindy Van Dover knows the submersible better than any other scientist. Now the director of the Duke University Marine Laboratory in Beaufort, North Carolina, Van Dover spent nearly three years training and working as an *Alvin* pilot after getting her doctorate at Woods Hole. She is the only researcher, and the only woman, to have piloted the craft, choosing to detour from a conventional research career because the job offered opportunities that few researchers can get. "I wanted to do a lot of dives," she says. "It was about wanting to see the environment that I was studying."

In 1991, when Van Dover gave up her pilot's seat and went back to being an occasional passenger in *Alvin*, one of the things she missed most was the view. The old sphere, built in 1973, had three 13-centimetre-wide viewports, one on each side for passengers and one at the front for the pilot. Through those saucer-sized side ports, Van Dover says, "sometimes it was a great view but often it wasn't, so you'd end up watching a video of what the pilot was doing". The new sphere retains the side windows but has three 18-centimetre-wide ones at the front, giving researchers a field of view that overlaps with the pilot's (see 'Voyager to the bottom of the sea').

The space inside the sphere has also grown, which should make for less-painful rides. With an internal diameter of less than 2 metres, the old sphere had no room for seats; scientists had to huddle on the floor, and lean awkwardly against the sloped walls. After a 2-hour descent to the bottom, limbs would stiffen up. "You switch your position because your legs are so uncomfortable and you rock the whole submersible," says Chris German, a marine geochemist at WHOI and chief scientist for the US National Deep Submergence Facility in Woods Hole, which coordinates the use of *Alvin*. "That can compromise taking the samples."

The new sphere has 18% more interior space and will have ergonomically designed seats. And while sampling or exploring, researchers will be able to take advantage of brighter lights and improved camera systems.

Scientists are keen to use the tools to explore the geology and biology of the mid-ocean ridges, the 60,000-kilometre-long chain of volcanic fissures that forge new ocean crust. So far, oceanographers have surveyed only a tiny fraction of the ridge system. The ridges and their flanks hide a subsurface plumbing network of fissures and pores so extensive that it circulates a volume of water equal to the entire ocean about every million years or so, altering the chemistry of the seas. Future dives will also target the communities of microbes living beneath the sea floor, which has been estimated to constitute as much as one-third of the planet's biomass. And marine scientists want to study the vast abyssal

plains, which cover more than half of Earth's surface and are poorly explored; most of the species that researchers pull up in samples from abyssal sites are new to science.

"WHAT IS THE FUTURE OF DEEP SUBMERGENCE RESEARCH?"

But some places that scientists want to visit lie deeper than 4,500 metres and so will remain off-limits to *Alvin* for now. The National Science Foundation, which is funding most of the *Alvin* redesign, could not afford to build a 6,500-metre-class sub all at once. In particular, such deep dives require more energy and it would be prohibitively expensive to build a submersible that uses current battery technology to dive frequently and to greater depths.

Rich Lutz, an oceanographer at Rutgers University in New Brunswick, New Jersey, who has racked up more *Alvin* dives than perhaps any scientist, was surprised to learn that the current overhaul won't add to the sub's depth range. "I certainly would have assumed that one of those improvements would be the capability of going down to 6,500 metres," he says.

BOTTOM DWELLER

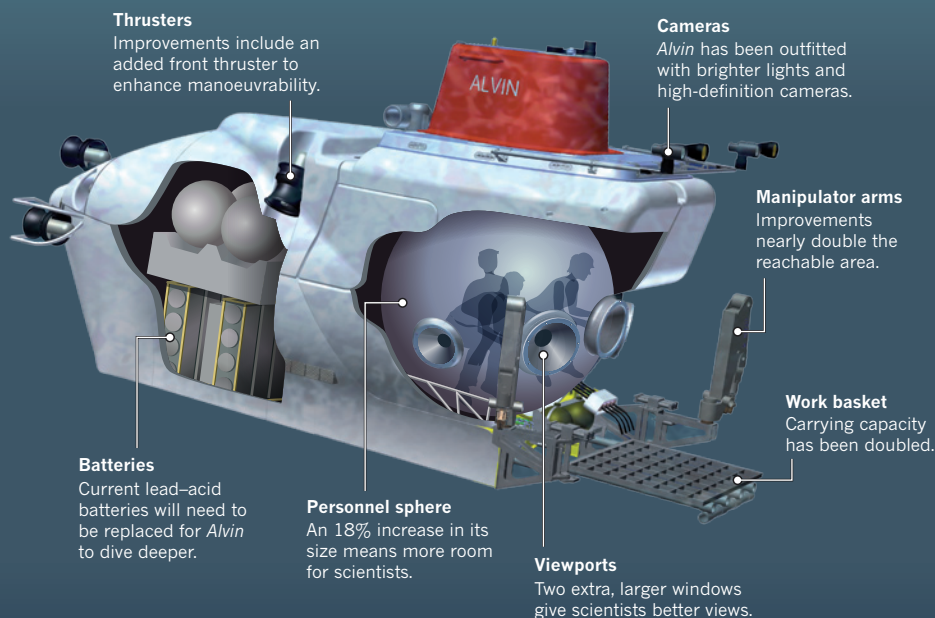
Yet depth isn't everything. Despite the excitement over record-breaking dives by China's *Jiaolong* (7,020 metres) and Cameron's *Deepsea Challenger* (11,000 metres), researchers don't go deep very often. Fewer than 10% of the dives by the French and Japanese submersibles approach their limits. Vincent Rigaud, manager of the underwater systems department at the French Research Institute for Exploration of the Sea in La Seyne-sur-Mer, says that research is increasingly focusing on shallower depths. "The science today is linked to [minerals] exploitation or areas where there is an impact from man's activities or man's interests," he says.

Some US researchers are more upset about other limitations of the new *Alvin*. "It's a superb tool for what it does but it could have been a hell of a lot better," says Bruce Robison, a marine ecologist at the Monterey Bay Aquarium Research Institute in Moss Landing, California. In 2004, Robison was a member of a committee of the National Research Council that assessed the future needs of deep submergence science. One of his biggest wishes was for a human-occupied vehicle (HOV) that could adjust its buoyancy to stop at any depth, allowing researchers to work well above the ocean floor. Most of the biological and chemical process in the ocean take place in the water column, not on the sea floor, Robison says. Yet neither the old *Alvin* nor the rebuilt sub can adjust its buoyancy enough to linger above the sea floor without draining its batteries. So nearly all trips go straight to the bottom.

Along with concerns over *Alvin's* limitations, US researchers also lament the loss of several scientific submersibles. Funding woes forced Florida International University's Harbor Branch Oceanographic Institute in Fort Pierce to retire its two submersibles, the Johnson Sea Links, last year. And the National Oceanic and Atmospheric Administration plans to stop funding the National Undersea Research Program, which supports two *Pisces* submersibles located in Hawaii. The two HOVs are known for their work exploring seamounts in the Pacific Ocean, including the newest Hawaiian volcano. Robison bemoans the loss of

VOYAGER TO THE BOTTOM OF THE SEA

A US\$40-million upgrade to *Alvin* will lead to noticeable improvements for researchers, but the submersible will have the same depth limit, 4,500 metres, as the previous one. Only five nations — and entrepreneur James Cameron — have research submersibles that can go deeper than 4,000 metres.



ALVIN ILLUSTRATION BY MEGAN CARROLL. © WOODS HOLE OCEANOGRAPHIC INSTITUTION

all the other US scientific submersibles. “As good as *Alvin* is,” he says “that isn’t enough.”

Veteran explorer Robert Ballard is not shedding any tears over the HOVs. “I consider that a dead paradigm,” says the marine geologist who discovered the *Titanic* and co-led the cruise that discovered the first hydrothermal vents; he used *Alvin* to explore them both. He now relies on autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs), which can be controlled by researchers on a ship or from halfway across the world. “AUVs augmented with ROVs is where I see the future. I don’t need any people in them.”

Alvin is limited to 4–5 hours of work at depth, whereas ROVs can spend days on the bottom and some AUVs can explore for months. And unmanned vehicles don’t require such large support ships and crews as HOVs, which can trim the cost of a mission substantially. Thanks to their data links, ROVs also enable essentially unlimited numbers of researchers around the world to take part in missions.

The ROV Jason 2 has become a favourite tool of US marine geologists, who love its flexibility in exploring and collecting samples down to 6,500 metres. But its heavy steel cable, which supplies power and transmits data, prevents the ROV from diving deeper and complicates missions.

Just up the hill from where *Alvin* is taking shape at Woods Hole, researchers are working on an experimental vehicle that will sever that troublesome umbilical cord while retaining the advantages of an ROV. Canary yellow and the size of a delivery van, Nereus made headlines in 2009 by reaching the bottom of Challenger Deep in the South Pacific — at 11,000 metres, the lowest spot on Earth. Nereus can operate either autonomously or connected to a surface ship through a fibre-optic cable.

But the tether, just three times the thickness of a human hair, breaks easily. So the Woods Hole team is now working to cut the physical link with the surface. This month, off Guam, the crew will test a laser system on Nereus that sends the data

to a receiver hanging down from a ship. Researchers will control Nereus using acoustic signals from the ship.

Despite the advantages of ROVs, most scientists say that there is no substitute for exploring the depths in person. “There is still a place for having a human brain and eyes at the bottom of the ocean,” says Humphris. To those who say there is no need for HOVs, she says “It’s a little bit like saying, ‘Why would you go to the Grand Canyon rather than watching a movie about it?’”

Alvin is also one of the best recruiting tools in marine science. “It’s what has brought the very best minds to the deep-sea oceanographic community,” says Lutz. “You’re able to tell them, ‘You come work with us and we’re going to take you a mile and a half down to the bottom of the sea in a submersible,’ as opposed to saying, ‘We’re going to take you out on a ship and you’re going to watch it on a television screen.’”

Private money could help to preserve some opportunities for deep-sea science. Eric Schmidt, former chief executive of Google, purchased an 83-metre-long ship in 2010 that he retrofitted for oceanographic science. His Schmidt Ocean Institute will start research cruises with the vessel next year and will test existing ROVs with the aim of purchasing one in a couple of years and making it available for researchers.

And Cameron’s dive aboard his one-person *Deepsea Challenger* fired up the imagination of researchers such as Lutz. “You look at it as a scientist and say, ‘Shit! I’d go down in that thing.’”

Cameron says that his sub is available for anyone willing to put up the money for further expeditions. “I’m more than happy to put in the sub and the technology. It’s not doing any good sitting in my barn,” he says.

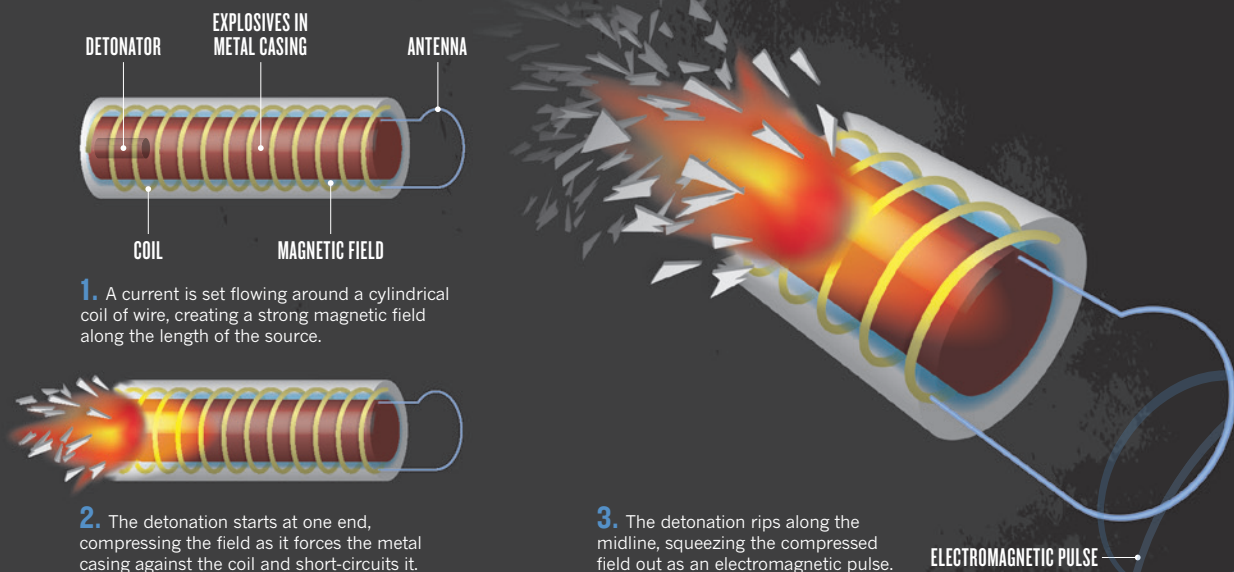
In the meantime, the hopes of the US oceanographic community ride on *Alvin*, which currently rests in pieces at Woods Hole. Looking up at the gleaming passenger sphere, Strickrott takes a moment to picture the day his crew has completed all its work and *Alvin* is ready to dive once again. “When it gets rolled out here, it’s spotless,” he says. “It’s a beautiful thing.” ■

Richard Monastersky is a features editor for Nature.

➔ **NATURE.COM**
For an interview with
filmmaker James
Cameron visit:
go.nature.com/uvjdyw

E-BLAST

The simplest kind of electromagnetic-pulse source works by blowing itself up.



WASTED ENERGY

DESPITE 50 YEARS OF RESEARCH ON HIGH-POWER MICROWAVES, THE US MILITARY HAS YET TO PRODUCE A USABLE WEAPON.

BY SHARON WEINBERGER

For some Pentagon officials, the demonstration in October 2007 must have seemed like a dream come true — an opportunity to blast reporters with a beam of energy that causes searing pain.

The event in Quantico, Virginia, was to be a rare public showing for the US Air Force's Active Denial System: a prototype non-lethal crowd-control weapon that emits a beam of microwaves at 95 gigahertz. Radiation at that frequency penetrates less than half a millimetre into the skin, so the beam was supposed to deliver an intense burning sensation to anyone in its path, forcing them to move away, but without, in theory, causing permanent damage.

However, the day of the test was cold and

rainy. The water droplets in the air did what moisture always does: they absorbed the microwaves. And when some of the reporters volunteered to expose themselves to the attenuated beam, they found that on such a raw day, the warmth was very pleasant.

A demonstration of the system on a sunny day this March proved more successful. But that hasn't changed a fundamental reality for the Pentagon's only acknowledged, fully developed high-power microwave (HPM) weapon: no one seems to want it. Although the Active Denial System works (mostly) as advertised, its massive size, energy consumption and technical complexity make it effectively unusable on the battlefield.

The story is much the same in other areas of HPM weapons development, which began as an East–West technology race nearly 50 years ago. In the United States, where spending on electromagnetic weapons is down from cold-war levels, but remains at some US\$47 million per year, progress is elusive. “There’s lots of smoke and mirrors,” says Peter Zimmerman, an emeritus nuclear physicist at King’s College London and former chief scientist of the US Arms Control and Disarmament Agency in Washington DC. Although future research may yield scientific progress, he adds, “I cannot see they will build a useful, deployable weapon”.

For many critics, the US HPM programme has become a study in wishful thinking, exacerbated by a culture of secrecy that makes real progress even more difficult.

The quest to build an electromagnetic weapon — an e-bomb, in military jargon — was sparked on 8 July 1962, when the United States carried out Starfish Prime, the largest high-altitude nuclear test that had ever been attempted. The 1.4-megaton thermonuclear warhead, detonated 400 kilometres above the central Pacific Ocean at 9 seconds past 11 p.m., Hawaii time, blasted huge swarms of charged particles outwards along Earth’s magnetic field. Their gyrations generated a pulse of microwave energy that drove measuring instruments off the scale. Artificial auroras lit up the night across swathes of ocean. And in Honolulu, more than 1,300 kilometres from the detonation point, the pulse set off burglar alarms, knocked out street lights and tripped power-line circuit breakers.

Nothing like Starfish Prime has been seen since August 1963, when the Partial Test Ban Treaty outlawed nuclear explosions anywhere but underground. But the test showcased the potential destructiveness of an electromagnetic pulse to military planners on both sides of the cold-war divide, and launched them into a race to harness it as a weapon using a non-nuclear source.

POWER CUT

The US Air Force has been the main funder of the country’s HPM programme from the beginning. At first, its goal was a weapon capable of taking out an enemy’s computers, communication systems and other electronics. In theory, the idea remains compelling: an e-bomb would be able to fire microwave ‘bullets’ at the speed of light and, if tuned to the right frequencies, disable its targets without collateral damage. Cars could be stopped in their tracks, radars blinded and computers destroyed, with no need for high explosives.

But that goal has foundered on the HPM weapon’s main technical challenge: generating a pulse that is directed enough to pick out a specific target and powerful enough to have an effect when it gets there, ideally using a generator that is small and light enough for an aeroplane or missile to lift.

A battery-powered device can generate an HPM pulse, but producing the kind of highly concentrated power needed to destroy electronics typically requires detonating a conventional explosive inside a device that destroys itself in the act of pulsing (see ‘E-blast’). Because doing this inside a piloted aircraft is risky — “a few pounds in the right place will take down anything”, notes Zimmerman — the Air Force has in recent years pursued HPM weapons designed for single-use missiles.

For example, the Counter-electronics High-power Microwave Advanced Missile Project (CHAMP) is an experimental cruise missile designed to take out electronic targets such as production sites for weapons of mass destruction. Neither the Air Force nor Boeing, its main contractor for CHAMP, will discuss technical details of the programme. But the project is just

“I’D RATHER TERRORISTS SPENT ALL THEIR TIME WORKING ON AN HPM WEAPON THAN CAR BOMBS.”

a prototype; when CHAMP was flight-tested last year, it still didn’t include the HPM payload.

It is possible to make a microwave generator compact enough for a missile. Engineers at Texas Tech University in Lubbock have developed an experimental explosive-based source less than 2 metres long and 16 centimetres in diameter (M. A. Elsayed *et al. Rev. Sci. Instrum.* 83, 024705; 2012). But lead developer Andreas Neuber points out that there are physical limits: to maximize the microwave power while keeping the system small, the engineers had to increase the internal electrical field. The result can be a catastrophic failure of the system’s insulating materials that short-circuits it before the system can build up much power.

Even if the military succeeds in packaging an HPM system, there is serious doubt over how effective the pulses will be when they hit their targets. In the late 1980s, a device called Gypsy successfully took out a bank of personal computers during the Air Force’s first unclassified test of a microwave weapon. But building on that success “became an incredibly difficult research project”, says Doug Beason, a physicist who was associate director for threat reduction at the Los Alamos National Laboratory in New Mexico until 2008, and wrote *The E-Bomb* (Da Capo, 2005), a discussion of directed-energy weapons. “You could understand how microwaves affected components of electronic circuits — transistors, capacitors, inductors and all that. But when you started putting them together in complex circuits, it became more of a stochastic process and you wouldn’t always get the same results each time.”

There is similar uncertainty over how electromagnetic energy flows through enclosures such as buildings. The process is chaotic, says Edl Schamiloglu, an electrical engineer at

the University of New Mexico in Albuquerque who is involved in a multi-university research initiative funded by the US defence department to improve such predictions. “When an electromagnetic ray or wave-beam enters the enclosure,” he says, “it will continue bouncing around and not repeat its trajectory.”

In short, more than 20 years after the Gypsy test, scientists still can’t reliably predict the damage a weapon would do. And that is without even considering the countermeasures that an adversary might use, which could be as elementary as surrounding sensitive electronics with a Faraday cage — the equivalent of the aluminium mesh used to shield microwave ovens.

The effort to disable electronics has remained mostly secret. But in 2001, the Air Force publicly announced that it had made substantial progress in developing microwave

weapons that target people, when it unveiled the Active Denial System.

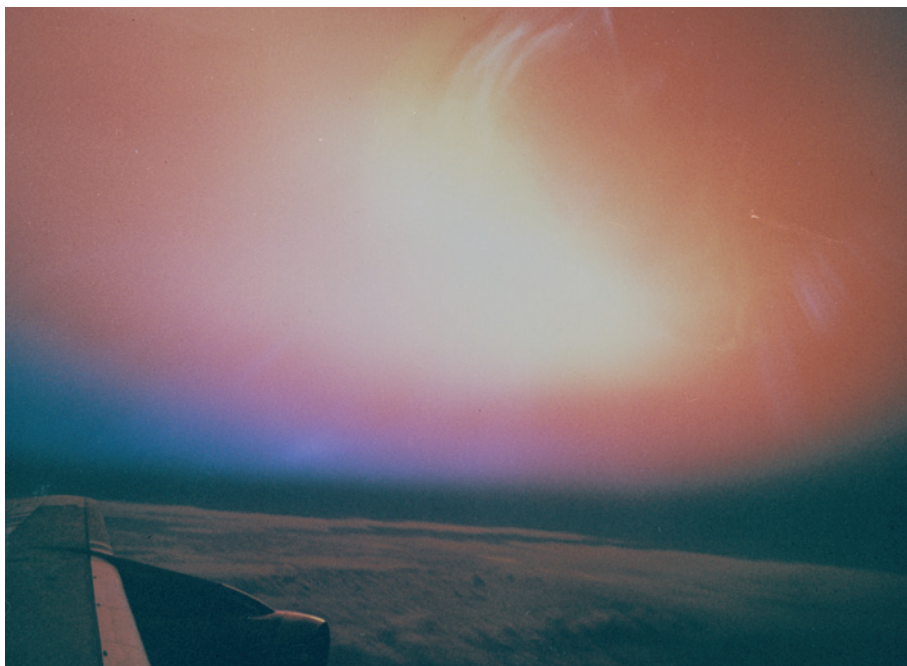
Development of the system began in the 1990s with the Air Force’s efforts to explore the biological effects of microwaves. A project code-named Hello studied how to modulate the clicking or buzzing sounds produced by microwave heating in the inner ear, to produce psychologically devastating ‘voices in the head’. ‘Goodbye’ explored the use of microwaves for crowd control. And ‘Good Night’ looked at whether they could be used to kill people.

HELLO GOODBYE

Only the Goodbye effect went into development as a weapon. Further bioeffects research was conducted in secrecy at Brooks Air Force Base near San Antonio in Texas, but even that programme almost stalled when the weapon was ready to move from animal to human testing. Hans Mark, a nuclear engineer at the University of Texas at Austin who was then the Pentagon’s director of defence research and engineering, paid a visit to Brooks in 2000 to check out the work. “Dr Mark didn’t believe in the effect,” recalls Beason, “and he actually had a shouting match with one of the main researchers.” But Mark’s approval was needed to advance the project, so he agreed to be subjected to the beam.

The Air Force got its human tests. The Brooks scientists joke that “you’ve never seen a political appointee run so fast”, says Beason.

Mark says that his doubts about the Goodbye effect were rooted in what he calls the “extravagant claims” made by its advocates. If nothing else, he says, the superconducting electromagnet that powered the system’s pulse generator required a cooling system too big and cumbersome to be used in the



The 1962 Starfish Prime nuclear test set off an electromagnetic pulse that sparked a weapons programme.

field. Mark says that he allowed the system to proceed to human testing not because he was convinced that it would work, but because after exposing himself to the beam, he decided that human testing at least wouldn't harm anyone. "Almost all of this programme has been a waste of money," he says.

Mark's concerns have proved prescient: efforts to deploy the weapon have been futile. At the 2001 unveiling, the defence department touted the Active Denial System for use in peacekeeping missions in places such as Kosovo and Somalia. But after the invasion of Iraq in 2003, when the US Joint Non-Lethal Weapons Directorate offered to deploy the Active Denial System to the region, it was rebuffed.

"We knew it wasn't reliable," said Franz Gayl, the Marine Corps's science and technology adviser, in an interview last year. Worse, he said, the pulse generator was so big that it had to be carried on its own utility vehicle. "That was a recipe for disaster," said Gayl, "because the operators are going to be a target." And worst of all, he said, before use the system had to be cooled down to 4 kelvin — a process that took 16 hours.

The defence department tried to deploy the weapon in Afghanistan in 2010, but it was sent home unused. In the same year, California rejected a smaller version meant for use in prisons. The device was built by defence contractor Raytheon of Waltham, Massachusetts, which declines to discuss it.

Other weapons have fared little better. The Air Force Research Laboratory developed an HPM system called MAXPOWER to detonate roadside bombs remotely, but it was the size of an articulated lorry — too unwieldy to be deployed in Afghanistan. The Joint Improvised Explosive Device Defeat Organization, the defence department's bomb-fighting agency,

declined to discuss the system, citing classification issues. But it did say that, as of 2011, it was not funding MAXPOWER.

In July, General Norton Schwartz, the Air Force chief who retired last month, warned that the service would have to withdraw from some science efforts amid budgets cuts, but that HPM technology would still be pursued. It "clearly has potential", he told the trade magazine *Aviation Week & Space Technology*, warning that countries such as Russia could be ahead of the United States.

THE MICROWAVE GAP

The concern that other nations, or even terrorists, could be working on similar technology seems to have been one of the prime motivations for the US military to continue investing in microwave weaponry, despite the apparent lack of progress. According to a 2009 briefing on non-lethal technologies prepared by the Office of Naval Research and obtained under the Freedom of Information Act, Russia, China and even Iran are pursuing HPM programmes — and the UK Defence Science and Technology Laboratory at Fort Halstead is sponsoring a classified car-stopping programme.

But such programmes are not necessarily proof that the cold-war HPM arms race is still going on. At least some countries may — like the United States — be conducting research out of fear of becoming vulnerable to such weapons. Modern technologies such as mobile phones are particularly susceptible to HPMS, says Michael Suhrke, head of the electromagnetic effects and threats business unit at the Fraunhofer Institute for Technological Trend

Analysis in Euskirchen, Germany.

As for HPM weapons in the hands of terrorists, many scientists regard that threat as far-fetched at best. Even if terrorist groups had the sophistication to carry out the necessary testing, says Yousaf Butt, a physicist in the high-energy astrophysics division at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, why would they? A microwave weapon of any magnitude would probably have to be powered by explosives. And if they had that kind of material, he says, "why wouldn't they just explode it?"

"Is it conceivable?" asks Philip Coyle, who in 2010–11 served as associate director for national security and international affairs in the White House Office of Science and Technology Policy, and is now a senior fellow at the Center for Arms Control and Non-Proliferation, a think tank based in Washington DC. "Barely, I think. I wouldn't take it for granted that terrorists couldn't do it. But I'd rather terrorists spent all their time working on [an HPM weapon] than car bombs."

Experts still disagree on whether HPMS might eventually make useful weapons. But one thing is clear: the mythical e-bomb capable of stopping cars or planes has not yet materialized on the battlefield. Asked whether the Air Force had produced any operational weapons, its research lab said only: "Due to operational concerns, we are unable to respond to this question."

The secrecy that surrounds HPM weapons research seems to have greatly exacerbated technical obstacles to the programme. In 2007, for example, a report on directed-energy weapons by the Defense Science Board said that the Pentagon had not effectively used data collected by university researchers to understand microwave effects. The Air Force claims that sharing is better now. But working in a field shrouded in secrecy still affects how information is disseminated. Neuber, for example, could agree to answer questions for this article only if he replied in writing, and only after his responses had been cleared through the US Army office that sponsors his team's work.

"Working in an area that is to a large extent of military interest requires playing by a set of different rules to some extent," he wrote. "Some flow of information is not as free as in other areas of the research endeavour."

To John Alexander, a retired army colonel who once headed the non-lethal weapons programme at Los Alamos National Laboratory, the secrecy reinforces the air of fantasy around the whole endeavour. "My point is always: chemistry and physics work the same way for everyone, and there are smart folks out there, so who are you trying to fool?" he says. "The people not getting adequate information were our own commanders." ■ [SEE EDITORIAL P.177](#)

Sharon Weinberger is a freelance writer in Washington DC.

COMMENT

FUNDING In defence of more scrutiny for elite grant-holders at the NIH **p.203**

NOVELS Celebrating the acerbic 'radium age' of science fiction **p.204**



CLIMATE Glacier photographer captures vanishing frozen landscapes **p.206**

ART Information designer Edward Tufte sculpts Feynman equations **p.207**

ILLUSTRATION BY DAVID PARKINS



Predicting scientific success

Daniel E. Acuna, Stefano Allesina and Konrad P. Kording present a formula to estimate the future h -index of life scientists.

We research scientists often worry about the future of our careers. Is our research an exciting path or a dead end that will end our careers prematurely? Predicting scientific trajectories is a daily task for hiring committees, funding agencies and department heads who probe CVs searching for signs of scientific potential.

One popular measure of success is physicist Jorge Hirsch's h -index¹, which captures the quality (citations) and quantity (number) of papers, thus representing scientific achievements better than either factor alone. A scientist has an h -index of n if he or she has published n articles receiving at least n citations each². Einstein, Darwin and Feynman, for example, have impressive h -indices of 96,

63 and 53, respectively. According to Hirsch, an h -index of 12 for a physicist — meaning 12 papers with at least 12 citations each — could qualify him or her for tenure at a major university.

However, the h -index³ and similar metrics⁴ can capture only past accomplishments, not future achievements⁵. Here we attempt to predict the future h -index of scientists on the basis of features found in most CVs.

We maintain that the best way of predicting a scientist's future success is for peers to evaluate scientific contributions and research depth, but think that our methods could be valuable complementary tools.

The typical research CV contains information on the number of publications,

those in high-profile journals, the h -index and collaborators. One can also infer interdisciplinary breadth, the length and quality of training, the amount of funding received and even the standing of the scientist's PhD adviser. Such factors are taken into account for hiring decisions, but how should they be weighted? Fortunately, obtaining data on the scientific activities of individual researchers has never been easier. Using all of these features, we can begin to probe the scientific enterprise statistically.

VITAL STATISTICS

To construct a formula to predict future h -index, we assembled a large data set and analysed it using machine-learning techniques. Our initial sample from academic-tree.org — a crowd-sourced website listing scientists' mentors, trainees and collaborators — contains the names and institutions of about 34,800 neuroscientists, 2,000 scientists studying the fruitfly *Drosophila* and 1,300 evolutionary researchers. We matched these authors to records in Scopus, an online database of academic papers and citation data. We restricted our analysis to authors who had accrued an h -index greater than 4 (to exclude inactive scientists); to publications after 1995 (because electronic records are sparse before then); to authors who had published their first manuscript in the past 5–12 years; and to authors who were identifiable in Scopus.

That left us with 3,085 neuroscientists, 57 *Drosophila* researchers and 151 evolutionary scientists for whom we constructed a history of publication, citation and funding.

For each year since the first article published by a given scientist, we used the features that were available at the time to forecast their h -index a number of years into the future. For example, we reconstructed how the CV features of a scientist looked five years after publishing his or her first article, and found a relationship between those features and the reconstructed h -index five years on.

Starting with neuroscientists, we attempted to predict the h -index of each scientist 5 years ahead — a timescale relevant for tenure decisions — using a linear

NATURE.COM

For more on science metrics, see: go.nature.com/nj2xqk

regression with elastic net regularization⁶ (see Supplementary Information at go.nature.com/mtvuzr). The ▶

METRICS

Predict your future *h*-index

These are approximate equations for predicting the *h*-index of neuroscientists in the future. They are probably reasonably

precise for life scientists, but likely to be less meaningful for the other sciences. Try it for yourself online at go.nature.com/z4rrroc.

● Predicting next year ($R^2=0.92$):

$$h_{+1} = 0.76 + 0.37\sqrt{n} + 0.97h - 0.07y + 0.02j + 0.03q$$

● Predicting 5 years into the future ($R^2=0.67$):

$$h_{+5} = 4 + 1.58\sqrt{n} + 0.86h - 0.35y + 0.06j + 0.2q$$

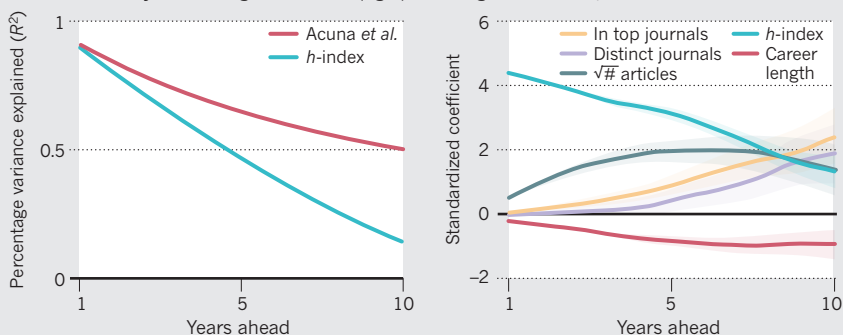
● Predicting 10 years into the future ($R^2=0.48$):

$$h_{+10} = 8.73 + 1.33\sqrt{n} + 0.48h - 0.41y + 0.52j + 0.82q$$

Key: *n*, number of articles written; *h*, current *h*-index; *y*, years since publishing first article; *j*, number of distinct journals published in; *q*, number of articles in *Nature*, *Science*, *Nature Neuroscience*, *Proceedings of the National Academy of Sciences* and *Neuron*.

PATHS TO SUCCESS

The accuracy of future *h*-index prediction decreases over time, but the Acuna *et al.* formula predicts future *h*-index better than does current *h*-index alone (left). The contribution of each factor to the formula accuracy also changes over time (right). Shading indicates 95% confidence error bars.



► model predicted the future *h*-index accurately, yielding a respectable $R^2=0.67$, cross-validated across scientists (an R^2 of 1 would imply that the model predicts the data perfectly). A simplified model containing only the number of published articles, the *h*-index, years since first publication, number of publications in prestigious neuroscience journals (*Nature*, *Science*, *Nature Neuroscience*, *Neuron* and the *Proceedings of the National Academy of Sciences*) and the number of distinct journals still performed nearly equally well ($R^2=0.66$; see 'Predict your future *h*-index').

Predicting the future careers of *Drosophila* and evolutionary scientists leads to somewhat worse predictions ($R^2=0.54$ and $R^2=0.61$, respectively, based on scientists 3–15 years into their careers) but still better than predictions based on the *h*-index alone ($R^2=0.38$ and $R^2=0.39$, respectively). This indicates that generalizations to other fields within and outside of life science may be limited¹. But for neuroscientists, at least, the predictions extend well to longer periods of time, such as ten years into the future

($R^2=0.52$). Over time, using just the *h*-index performs much worse than taking all features into account (see 'Paths to success', left panel).

The main five predictive features change in importance for predicting *h*-indices over increasingly longer periods (see 'Paths to success', right panel). The power of the *h*-index declines. The number of articles written, the diversity of publication in distinct journals and the number of articles published in five prestigious journals all become increasingly influential over time.

FUTURE FORTUNES

It is risky to make any causal interpretations of these results. However, we will briefly speculate on why these features might be important predictors of future success. Some features directly affect the potential for a high *h*-index, such as the number of articles written. These features can also indirectly affect a scientist's future success, because scientists who are productive and publish many papers tend to remain productive. Publishing in many different journals may lead to fewer overlapping populations of scientists

who cite the work, and hence higher growth potential for articles. A scientist who has published in several distinct journals is also likely to be someone with broad training who contributes in many ways. The number of publications in leading journals can increase the visibility of a scientist's other papers, past and future.

If promotion, hiring or funding were largely based on indices (*h*-index, the model used here or any other measure), then some scientists would adapt their behaviour to maximize their chances of success. Models such as ours that take into account several dimensions of scientific careers should be more difficult for researchers to game than those that focus on a single measure.

Our formula is particularly useful for funding agencies, peer reviewers and hiring committees who have to deal with vast numbers of applications and can give each only a cursory examination. Statistical techniques have the advantage of returning results instantaneously and in an unbiased way. Building and analysing massive data sets to track scientific careers could also help to identify potential gender, racial and other biases^{7–9} and advance our understanding of how science develops.

Although our findings and predictions may not alleviate scientists' angst over their careers, the results offer some comfort by showing that the future is not so random. The occasional rejection of a paper may feel unjust and indiscriminate, but in the long run, such factors seem to average out, rendering *h*-index trajectories relatively predictable. ■

Daniel E. Acuna is a research associate at the Rehabilitation Institute of Chicago, Illinois 60611, USA, and a research affiliate in biomedical engineering at Northwestern University, Evanston, Illinois 60208, USA.

Stefano Allesina is assistant professor in ecology and evolution and at the Computation Institute at the University of Chicago, Illinois 60637, USA.

Konrad P. Kording is associate professor of physical medicine and rehabilitation, physiology, and applied mathematics at Northwestern University, and at the Rehabilitation Institute of Chicago. e-mail: daniel.acuna@northwestern.edu

1. Hirsch, J. E. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572 (2005).
2. Redner, S. *J. Stat. Mech. Theory Exp.* **3**, L03005 (2010).
3. Peterson, I. *ScienceNews* 2 December 2005; available at <http://go.nature.com/iawd5o>.
4. Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E. & Herrera, F. *J. Informetr.* **3**, 273–289 (2009).
5. Hirsch, J. E. *Proc. Natl Acad. Sci. USA* **104**, 19193–19198 (2007).
6. Zou, H. & Hastie, T. *J. Roy. Stat. Soc. B* **67**, 301–320 (2005).
7. Dwan, K. *et al. PLoS ONE* **3**, e3081 (2008).
8. Ginther, D. K. *et al. Science* **333**, 1015–1019 (2011).
9. Allesina, S. *PLoS ONE* **6**, e21160 (2011).

Well-funded investigators should receive extra scrutiny

An added layer of review for elite grant-holders upholds the mission of the National Institutes of Health, says **Jeremy M. Berg**.

The US National Institutes of Health (NIH) receives many more grant applications proposing outstanding scientific projects than its budget can support. The overall success rate for grants in fiscal year 2011 was 18% — a historic low.

Last month, the agency announced a policy that will apply to the next round of R01 grants. Applications from investigators who receive more than US\$1 million a year in funding from the NIH (not including 'indirect costs' that go to the investigator's university) will be subjected to extra review. The agency's advisory councils will examine how distinct the proposed project is from the investigator's other funded work. If it is a grant renewal, they will consider whether the project has been productive, and look at its value to the investigator's research programme and collaborations. This analysis will be provided to NIH programme staff, who will make funding recommendations.

The policy has been criticized for adding to the NIH's administrative burden, and for potentially penalizing the most productive scientists. However, I believe that it will complement peer review and help to determine the best possible portfolio of NIH-funded research, short and long term.

The National Institute of General Medical Sciences (NIGMS) in Bethesda, Maryland — where I was director between 2003 and 2011 — has, for around two decades, given extra scrutiny to applications from investigators receiving more than \$750,000 per year in costs from all funding sources, including the pending NIGMS application. This policy is not a cap; rather, it is a useful tool for managing taxpayer resources to achieve the goals of the NIGMS and the NIH. The scrutiny has influenced the distribution of resources, either by denying or reducing funding for well-off applicants, or providing grants on the condition that others will not be renewed. This frees up money for other investigators and allows those who are well-funded to change the course of their research. Assuming that the NIH policy can be implemented effectively, it is a step in the right direction.

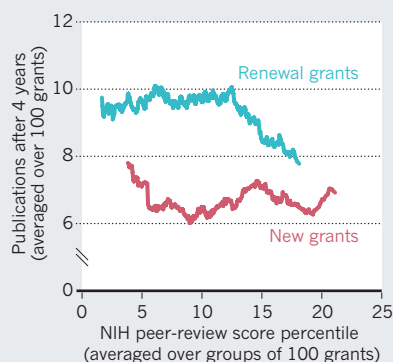
Critics of the NIH policy argue that although peer review is not perfect, it is the best system for evaluating scientific projects. Many reviewers try to ensure that a new project does not overlap significantly with one that is already funded. But they do not have

full access to NIH portfolio data, nor the time to analyse the information.

Peer review has several limitations. To compare and rank grant applications across hundreds of NIH sections, administrators assign each peer-review score a percentile value. The first percentile (the top 1%) corresponds to the most highly rated applications. These percentile values have uncertainties that stem from the potential biases of

SCORE DRAW

Higher grant scores do not imply more papers. Renewal grants show no significant differences up to the 13th percentile; new grants showed no significant correlation over the entire range.



primary reviewers, the characteristics of the process used to calculate percentiles, and the fact that an application is judged by whether it will be an important future scientific advance — which is inherently difficult to predict.

BANG FOR BUCK

During my time as director of the NIGMS, I analysed the power of peer-review scores to predict scientific productivity. This was judged by various measures, including numbers of publications, citations and highly cited publications four years after the grants were funded. The distribution of percentile scores for applications did correlate with future differences in productivity to some extent. But applications that were separated by five or even ten percentile points did not differ significantly in subsequent productivity (see 'Score draw'). Therefore, funding grants in strict percentile order does not necessarily mean that ensuing investigations will be of higher quality. So percentile score should not

be the only factor taken into consideration when making funding decisions.

There are good reasons to consider how much other funding an investigator is receiving. Projects in one lab almost always overlap, even if they are addressing different questions, so some of the running costs will already be covered. Giving another grant to the same investigator might have less impact than giving it to someone with nearly the same percentile score who has little or no other funding. In another NIGMS analysis, I showed that research productivity (measured by numbers of publications or citations and averaged over groups of investigators with similar levels of support) did not increase consistently as the level of support increased, but reached a plateau near \$700,000 in annual direct costs per investigator.

Some investigators are better than others at managing large amounts of resources. If an investigator is not performing as expected given the level of support, using the funds to support another researcher might be a better investment. The NIH needs to invest in a broad range of research; if scientists are well funded, their area could already be receiving enough support from other sources.

The layer of extra scrutiny is not ideal. I would prefer the limit to be less than \$1 million and to include funding outside the NIH. Exceptions to the policy also mean that it might be easy to dodge. For example, grants with multiple principal investigators are scrutinized only if all receive more than \$1 million in NIH funding, so I am concerned that some applicants may avoid extra review by adding a co-investigator who receives less.

I believe that special consideration should be given to investigators with strong proposals who have few or no other sources of funding, such as those at the beginning of their careers or established, productive investigators. Funding these applicants would probably have a bigger impact — by helping to develop a new lab or keeping an effective one functioning — rather than providing incremental support to an investigator who already has substantial other support. ■

Jeremy M. Berg is associate senior vice-chancellor for science strategy and planning (health sciences) at the University of Pittsburgh, Pennsylvania 15260, USA. e-mail: jberg@pitt.edu



Speculative fiction from 1904 to 1933 is often overshadowed by later 'golden age' work.

SCIENCE FICTION

The radium age

Joshua Glenn explores the dark, fascinating, largely forgotten science fiction of the early twentieth century.

We think we know science fiction. There were the 'scientific romance' years that stretched from around the mid-1860s to 1903, after which H. G. Wells lost his touch. And there was the 'golden age', from 1934 to the mid-1960s. But between those periods — and overshadowed by them — was a fascinating era that gave us such enduring memes as the berserk robot, the tyrannical superman and the sinister telepath.

I call that period, from 1904 to 1933, sci-fi's 'radium age'. It emerged when the speed of change in science and technology was inducing vertigo on both sides of the Atlantic. More cynical than its Victorian precursor yet less hard-boiled than the generation that followed, this is sci-fi offering a dizzying, visionary blend of acerbic social commentary and shock tactics. It yields telling insights into its context, the early twentieth century. Plus, it is fun to read.

➔ **NATURE.COM**
For more on science fiction, see the **Nature Web Focus:**
go.nature.com/mqc2jd

The scientific romances that preceded this era were strong on the fantastical but otherwise very different. The utopian

fantasies of Edward Bulwer-Lytton (*The Coming Race*, 1871), for instance, and the fantastic voyages by writers such as Jules Verne (*Journey to the Centre of the Earth*, 1864) and Edgar Allan Poe (*The Unparalleled Adventure of One Hans Pfaall*, 1835) were largely sentimental literature. They were by and for people who were inspired by US inventor Thomas Edison and believed that modern science and technology would improve the human condition and usher in the triumph

of millennia-old Christian values.

Radium-age writers were not out to beguile. They depict a human condition subverted or perverted by science and technology, not improved or redeemed. Aldous Huxley's 1932 *Brave New World*, with its devastating satire on corporate tyranny, behavioural conditioning and the advancement of biotechnology, is far from unique. Radium-age sci-fi tends towards the prophetic and uncanny, reflecting an era that saw the rise of nuclear physics and the revelation that the familiar — matter itself — is strange, even alien. The 1896 discovery of radioactivity, which led to the early twentieth-century insight that the atom is, at least in part, a state of energy, constantly in movement, is the perfect metaphor for an era in which life itself seemed out of control.

As demonstrated by Philipp Blom in *The Vertigo Years: Europe 1900–1914* (Basic Books, 2008), the actual experience of time and space was altered in this period by the proliferation of a range of stunning technological developments, from the telegraph and the telephone to aeroplanes and cinematic film. Thinking and perception were radically reoriented by developments in science, philosophy and the arts. These included Einstein's special theory of relativity; the work of phenomenological philosophers such as Edmund Husserl, who sought to objectively describe the subjective workings of consciousness; and Cubism, the artistic movement spearheaded by Pablo Picasso and Georges Braque that explored distortion and deconstruction. Social and cultural forms and norms, such as women's subordination to men — once regarded as natural, permanent 'elements' — were disintegrating.

During this astonishing period, sci-fi writers were obsessed with the future. The scientific romantics had conjured up simplistic utopias that remained firmly grounded in contemporary realities. By contrast, the radium-age novels, stories, movies and plays (see 'Key radium-age works') often lift off into previously unexplored realms.

Instead of offering progressive solutions

NOVEL VISIONS

Key radium-age works

Karel Čapek *R.U.R.* (1920)
Arthur Conan Doyle *The Lost World* (1912)
Hugo Gernsback *Ralph 124C 41+* (1911–12)
H. Rider Haggard *When the World Shook* (1919)
William Hope Hodgson *The Night Land* (1912)
Aldous Huxley *Brave New World* (1932)
Rudyard Kipling *With the Night Mail* (1905)

Jack London *The Scarlet Plague* (1912)
Charlotte Perkins Gilman *Herland* (1915)
Edward Shanks *The People of the Ruins* (1920)
George Bernard Shaw *Back to Methuselah* (1921)
Olaf Stapledon *Last and First Men* (1930)
Thea von Harbou *Metropolis* (1926)
Philip Wylie *Gladiator* (1930)
Yevgeny Zamyatin *WE* (1921)

to social unrest, writers satirized and exaggerated its causes and effects. The out-of-control robot can be read as a criticism of the efficiency-oriented theories of Frederick Winslow Taylor and the practices of industrialist Henry Ford. So *R.U.R.* (Czechoslovakian writer Karel Čapek's 1920 play, which introduced the term 'robot' to the English language) portrays mass production as alienating at best. In Jack London's post-apocalyptic *The Scarlet Plague* (1912), a race of barbarians descended from San Francisco's brutalized underclass roam the city's devastated remains after the fatal pandemic of 2133. And Charlotte Perkins Gilman's feminist novel *Herland* (1915) imagines an ideal community in which women aren't merely emancipated, but have done away with men altogether.

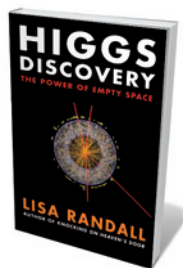
Politics are inevitably part of the mix — this is an era that encompassed the First World War, the Russian revolutions, and the rise of radical left- and right-wing movements. Many sci-fi authors, such as Čapek and Yevgeny Zamyatin, were leftists and liberals. But more conservative authors shared their utopianism and cynicism. For instance, Arthur Conan Doyle, inventor of Sherlock Holmes, wrote a series of ripping yarns starring a Professor Challenger, who discovers surviving dinosaurs, travels to Earth's "sensory cortex" and witnesses the end of life on the planet — all the while making the case for reconciling imagination and intuition with a sceptical scientific method.

Similarly, H. Rider Haggard, who had made his name with *King Solomon's Mines* (1885) and the Quartermain novels of the 1880s, created a disputatious trio in 1919's *When the World Shook* — an idealistic Anglican minister, a sardonic doctor and an adventurer whose world view hovers somewhere in between. Meanwhile, both the leftist London (in *The Scarlet Plague*) and the conservative English poet Edward Shanks (in *The People of the Ruins*) agree that the destruction of modern western society wouldn't be an entirely bad thing.

Fans of Philip K. Dick or Ursula K. Le Guin — writers belonging to what literary theorist Fredric Jameson has termed the "anti-anti-utopian" trend of the late 1960s and early 1970s — will find provocative antecedents here. Reading the dangerous visions of radium-age sci-fi, published in times as volatile as our own, destabilizes everything we take for granted. These books remind us that we need to regard our twenty-first-century forms and norms without sentimentality. ■

Joshua Glenn edits the blog *HiLobrow*, and co-founded *HiLoBooks* to reissue 'radium age' science fiction. *HiLoBooks'* edition of Arthur Conan Doyle's *The Poison Belt* is out now; H. Rider Haggard's *When the World Shook* is published in October. e-mail: jglenn@hilobrow.com

Books in brief



Higgs Discovery: The Power of Empty Space

Lisa Randall BODLEY HEAD 64 pp. £4.99 (2012)

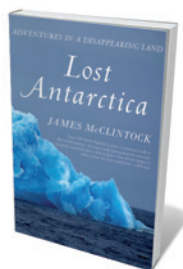
The Higgs bombshell on 4 July rocked the world of physics. In this slim volume, theoretical physicist Lisa Randall analyses the significance and implications of that momentous finding at Switzerland's Large Hadron Collider. She offers clear accounts of the Higgs mechanism and the role and modes of the Higgs's decay; follows the seven-month lead-up to the discovery; and speculates about what it all might mean for other areas of exploration, such as supersymmetry. A lucid, deft and engaging summation of dogged determination and "heroic engineering".



The Human Quest: Prospering Within Planetary Boundaries

Johan Rockström and Mattias Klum STOCKHOLM TEXT 314 pp. \$9.99 (2012)

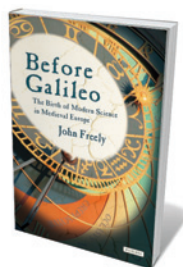
In 2009, Johan Rockström of the Stockholm Resilience Centre and his colleagues set out in *Nature* nine 'planetary boundaries' — numerical limits for processes that affect Earth's capacity to support human life, such as freshwater use and climate change. This lavishly illustrated e-book — with a foreword by former US president Bill Clinton and video clips from photographer Mattias Klum — extends this idea, laying out pressures and tipping points. Paramount among the changes needed, Rockström says, is a big shift in behaviour.



Lost Antarctica: Adventures in a Disappearing Land

James McClintock PALGRAVE MACMILLAN 256 pp. £16.99 (2012)

The sight of 50,000 king penguins on the Crozet Island Archipelago in the early 1980s sparked marine biologist James McClintock's fascination with Antarctic fauna. Now a veteran of the extreme south, McClintock shares the otherworldly wonders unveiled by decades of research. The book is packed with joys, from soft-coral 'trees' that replant themselves to the snoozing Weddell seal, stinking of putrid fish, that the author encountered in a dive hut. Running like a chill current through all is the climate-driven decimation of the ice on which these ecosystems depend.



Before Galileo: The Birth of Modern Science in Medieval Europe

John Freely OVERLOOK PRESS 352 pp. £18.51 (2012)

A thousand years before Galileo, the transmission of knowledge that survived the burning of the ancient Library of Alexandria began. Physicist John Freely traces this "tenuous Ariadne's thread" of classical learning that unspooled from Egypt through Byzantium and the Islamic world, finally emerging as Latin texts. Focusing on the trailblazers through this extraordinary millennium — from Bede, Averroës and al-Khwarizmi to Adelard of Bath, Robert Grosseteste and Roger Bacon — Freely ends with a coda on Copernicus, Kepler, Galileo himself and Newton. Shoulders of giants indeed.



The Scientists: An Epic of Discovery

edited by Andrew Robinson THAMES & HUDSON 304 pp. £24.95 (2012)

The human face of scientific breakthroughs from the sixteenth to the twentieth centuries is spotlighted in this sumptuously illustrated volume. Science writer Andrew Robinson, editing contributions from a stellar team of authors, groups 43 greats into six broad areas: Universe, Earth, Molecules and Matter, Inside the Atom, Life, and Body and Mind. From Alan Turing and Marie Curie to William Harvey and Chandrasekhar Venkata Raman, this is a sampler of the driven, complex, fascinating characters who fomented scientific revolutions.



J. BALOG/EXTREME ICE SURVEY

James Balog has developed new photographic equipment to monitor changes in glaciers such as the Ilulissat in Greenland.

Q&A James Balog

Archivist of ice

For six years, photographer James Balog has trained his lens on ice, capturing time-lapse images that have helped scientists to study how glaciers and ice sheets respond to climate conditions. With the documentary *Chasing Ice* soon to debut in US cinemas, Balog talks about the loss of landscapes.



Why are you interested in frozen landscapes?

When I was six years old, I had to walk home from school in a heavy snowstorm. It was a great contest between myself and nature. As a young man, I started ice climbing and mountaineering in the White Mountains in New Hampshire. When I looked at photographs of glaciers in Alaska I knew I wanted to spend time there.

When did you start photographing glaciers in time lapse?

After being blown away by the glacial retreat I saw while shooting a *National Geographic*

Portraits of Vanishing Glaciers

Rizzoli International: 2012. 288 pp. £29.95

Chasing Ice

Jeff Orlowski: 2012

cover story, 'The big thaw', in 2006. The next year I created the Extreme Ice Survey, deploying time-lapse cameras to monitor retreating ice in the Arctic and mountain areas, some so remote that our team was probably the first to visit them. Temperatures could reach -40°C and the winds 150 miles per hour. We had to tackle deep snow, torrential rain and falling rocks. Off-the-shelf gear wasn't robust enough, and it took six months of experimenting to come up with a reliable time-lapse system. Right now, we have 27 such cameras at 18 glaciers in Alaska, Greenland, Iceland, the Rocky

Mountains and at Mount Everest. Shooting every half-hour of daylight year-round, each one generates 8,500 frames per year. The footage provides scientists with information on the mechanics of glacial melting and gives the public evidence of how rapidly Earth's climate is changing.

What is the most dramatic moment you've caught?

Every year there are calving events in which ice falls off glaciers into the sea. The rate of ice loss in Greenland has doubled during the past 20 years, and this summer we've seen unprecedented rates of surface melting. We expected the Ilulissat Glacier on the west coast of Greenland to have a massive discharge of ice, so in summer 2008 two of my team members stood watch for weeks. They caught the biggest calving event ever recorded on film. It was as if the entire lower tip of Manhattan had broken off and fallen into the ocean, like an urban skyline toppling, with skyscraper-sized blocks of ice submerging and resurfacing.

What's the aim of the Extreme Ice Survey?

The problem of climate change is perception. Human brains and our economic system favour the status quo. We aim to collect powerful visual evidence so that people can understand what's really happening.

ILLUSTRATION: N. HIGGINS
PHOTO: J. ORLOWSKI/EXTREME ICE SURVEY

When satellites capture retreating glaciers from hundreds of kilometres away, the images may be beautiful but they're removed from human experience. We live on the ground. Retreating glaciers are where you can see climate change as it happens.

What have scientists learned from your images?

In 2006, on the Greenland Ice Sheet, I met glaciologists who said that glaciers are big lumbering things that respond to climate change on a scale of decades or centuries. We have learned that glaciers respond hourly to atmospheric conditions. Our visuals also suggest that subglacial floods may trigger some of the bigger calving events. We have nearly a million pictures in the archive now, although there's a lot of analysis yet to do.

How do you fund the ice survey?

A significant part of our initial funding was from the Expedition Council at *National Geographic*. Nikon gave us hardware. By partnering with scientists, we got help from the US National Science Foundation and NASA. But most of the funding over the past five years has been from private donors. It has been an absurdly difficult project to run without 'big science' government-scale funding and logistical support.

Do you have a science background?

For my master's in geomorphology, I researched Colorado's Big Thompson River flash flood of 1976. But I would not presume to call myself a scientist. When I was finishing my thesis, I remember looking at a stack of manila punchcards and deciding that I'd rather see the world through a camera than through data analysis. The data are incredibly important, but my calling is to understand the world through art.

What next?

The underlying theme of my work is the collision between human needs and nature. I'd like to do more with energy, in part because my grandfather died in a Pennsylvania coalmine collapse. We worked in the Gulf of Mexico after the Deepwater Horizon oil spill and are documenting changes in forest cover in the Rockies. And I am trying to make the ice survey financially stable so that the cameras can keep going for a long time. I feel a tremendous obligation to preserve a pictorial memory of these vanishing landscapes for the people of the future.

INTERVIEW BY JASCHA HOFFMAN

PHYSICS

Modelling Feynman

Daniel Cressey marvels at a gleaming depiction of the subatomic by the world's leading information designer.

Before the explosion in 'infographics' describing everything from cocktail mixology to US health-care spending, there was Edward Tufte.

Tufte, a statistician, is perhaps the world's most celebrated information designer. In *All Possible Photons* at his Manhattan gallery, ET Modern, Tufte has unleashed his love of artistic explanation in a series of sculptures instantly recognizable to anyone with a passing knowledge of particle physics. Minimalist 'graphics' consisting of stainless-steel tubing formed into straight and undulating lines, Tufte's rendering of Feynman diagrams transforms recondite scientific notation into abstract, glinting art. Tufte also plans to show versions more than five metres high at his sculpture park in Woodbury, Connecticut.

US physicist Richard Feynman created elegant tracers of lines, dots and arrows to describe and predict how subatomic particles interact. Feynman was not unaware of his diagrams' aesthetic appeal, and famously drove a van painted with a selection of them.

Tufte's matt or polished steel sculptures, mounted on the walls, are shorn of explanations as well as much of the detail that makes them scientifically useful, such as arrows and labels. Some are large and dominate their wall space. The most powerful artwork on display is the collection of 120 smaller pieces that give the show its name — a cluster representing all possible space and time paths of a particular interaction of photons. These form what Tufte

All Possible Photons: The Conceptual and Cognitive Art of Feynman Diagrams

ET Modern, New York

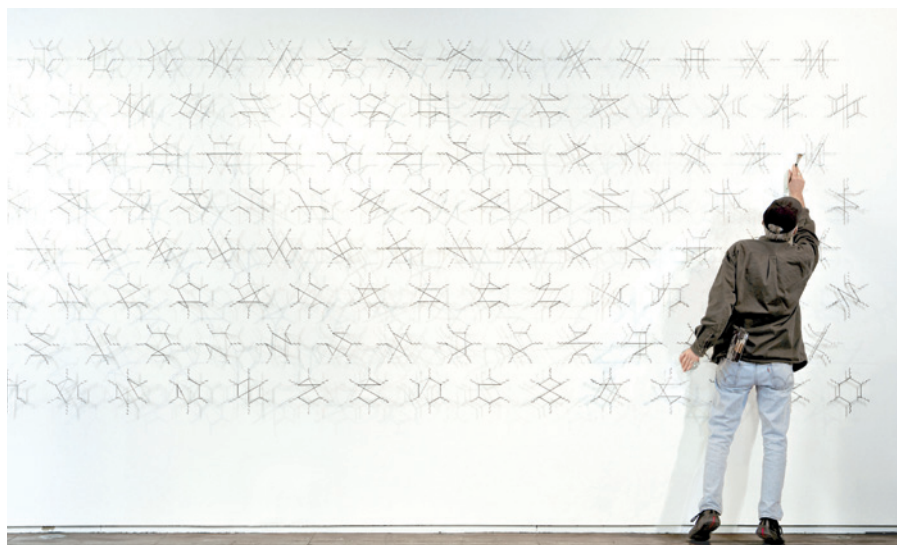
Until spring 2013; official opening 15 September.

calls "a complete enumeration of everything that could happen" in that instance.

This isn't the first time Tufte has ventured into Feynman territory. The diagrams feature as models of good design in Tufte's book *Beautiful Evidence* (Graphics, 2006). They are also referenced in one of his enormous *Rocket Science* sculptures of fantastical space probes attached to giant bazooka-style launchers. In the *Airstream Interplanetary Explorer* (2011), the probe is an iconic silvery Airstream caravan adorned with Feynman diagrams. Tufte's contention is that because subatomic particles everywhere in the Universe behave as shown by Feynman diagrams, these could work as communiqués to other worlds. As he has put it, "Better the cosmopolitan verbs of Nature's laws on spacecraft than the local proper nouns of national flags, earthly Gods and Goddesses, and government agency logos."

By focusing on the diagrams alone, the sculptures in *All Possible Photons* bring home the power of Feynman's achievement. There is beauty in his diagrams, but the real deal is what they can potentially describe — which is everything. ■

Daniel Cressey is a reporter for Nature.



Edward Tufte adjusts his *All Possible 6-Photon Scattering* (120 Space-Time Feynman Diagrams) (2012).

A. SEVERNY

Correspondence

Clarify the limits of climate models

Mark Maslin and Patrick Austin suggest that scientists should explain the statistical uncertainties in models of climate-change impacts (*Nature* **486**, 183–184; 2012), but these models are tools for insight, not prediction. A diversity of models benefits both science and science-based policy guidance.

We believe that future models will be more informative than those of today. The public-image problem of current models stems partly from scientists' failures to identify the limitations openly. It is important to distinguish between questions for which current models are useful as prediction engines and those for which the models merely probe possibilities. The role of science is to reflect on the plausibility and relevance of such possibilities.

Deep uncertainty is not foreign to policy-makers, who often have to weigh the advantages of deciding to wait against the potentially high costs of waiting.

David A. Stainforth,
Leonard A. Smith *London School of Economics, UK.*
d.a.stainforth@lse.ac.uk

ENCODE leads the way on big data

The ENCODE project offers a fresh perspective on big data by providing an organized framework for genomics (www.nature.com/encode). Other big-data efforts tend to focus on rapidly locating needles in petabyte-sized haystacks (such as finding the Higgs boson, for instance), whereas ENCODE aims to supply a structured overview.

ENCODE's organization of information is hierarchical, with raw data at the bottom and layers of annotation above. The processed summaries become progressively broader — for example, starting at the level of signals representing the degree to which DNA is bound by

transcription factors, moving on to the locations of sites where these factors bind, and then to overviews of regulatory networks. At the summit are the linked publications documenting the annotation.

The ENCODE data model could be useful in other fields: for example, astronomy and Earth science are in the process of organizing their reams of data (M. J. Raddick and A. S. Szalay *Science* **329**, 1028–1029; 2010), but don't yet compare with ENCODE in terms of the level of integration.

Mark Gerstein *Yale University, New Haven, Connecticut, USA.*
mark.gerstein@yale.edu

Animal variants in sexual behaviour

Andrew Barron and Mark Brown suggest that journalists, by merging animal and human behaviour, promote the myth that minority sexual preferences are an 'illness' (*Nature* **488**, 151–152; 2012). They include one of my articles among their examples — 'Bat bugs turn transsexual to avoid stabbing penises' (see go.nature.com/9iwdpb) — presumably because it uses 'transsexual' to describe insect behaviour. I contend that this term is valid shorthand for what the insects are doing, as well as catching the reader's attention.

Reporting of same-sex behaviour in animals can stray into troubling territory. In one example given by Barron and Brown, *The Sunday Times* wrote in 2006 that a study of male-oriented behaviour in sheep "could pave the way for breeding out homosexuality in humans". However, the problem here is not the terms used to describe animals' sex lives, but the leverage of the research paper to make a political point.

Homosexual behaviour is condemned as unnatural in some places around the world. By describing the huge variation in sexual behaviour across the

animal kingdom, the media may be helping to destroy forever the idea that gay, lesbian, bisexual and transgender people represent some kind of aberration.

Rowan Hooper *London, UK.*
rowan.hooper@newscientist.com

A code of conduct for European scientists

Progress in safeguarding research integrity is lamentably slow in Europe compared with the United States (*Nature* **488**, 7; 2012). Many European universities are still unsure of their responsibilities in policing research conduct.

Punishing misconduct makes no sense without prevention, given that one study found that up to one-third of all researchers engage in questionable research practices (D. Fanelli *PLoS ONE* **4**, e5738; 2009). What is needed is for each research institution to provide a detailed standard of responsible research conduct, along with mandatory education, so that researchers know how to uphold this standard in day-to-day practice.

We suggest that a European forum should be set up to educate researchers and their mentors in all matters relating to this standard. The forum would develop educational material, train and accredit teachers and remove discrepancies between research institutions.

Every university would need to set up a 'research integrity unit' to guide and supervise its research activities, and to ensure compliance with international standards.

We urge public and private research funders to lend their support to this venture, which would help to free European research from questionable research practices and make life harder for fraudsters.

Nils Axelsen *Statens Serum Institut, Copenhagen, Denmark.*
Xavier Bosch *Department of Internal Medicine, Hospital Clinic, University of Barcelona, Spain.*
xavbosch@clinic.ub.es

Wallace pioneered astrobiology too

The field of astrobiology, called 'exobiology' during the space programmes of the 1960s (*Nature* **488**, 160; 2012), was anticipated six decades earlier by Alfred Russel Wallace, better known as the co-discoverer of the principle of natural selection.

Next year is the centenary of Wallace's death, when his seminal contribution to biogeography and evolution will be celebrated. But he should also be remembered as a pioneer in astrobiology, whose hypotheses are still alive today.

Wallace introduced the concept of "astro-biology" in his popular book *Man's Place in the Universe* (Chapman & Hall, 1903). He reviewed the physical conditions required for organic life in terrestrial ecosystems and concluded that Earth is the only habitable planet in the Solar System.

He later published a monograph (*Is Mars Habitable?* Macmillan, 1907) evaluating astronomer Percival Lowell's suggestion that Mars could be "inhabited by a race of highly intelligent beings" (see *Nature* **74**, 587–589; 1906). After analysing what was then known about the Martian climate, temperature, possible presence of water and the 'canals' thought by Lowell to indicate intelligent life, Wallace roundly refuted this idea.

U. Kutschera *Institute of Biology, University of Kassel, Germany.*
kut@uni-kassel.de

CORRECTION

The obituary of Andrew Fielding Huxley (*Nature* **486**, 474; 2012) omitted to mention that Hugh Huxley and Jean Hanson also proposed a sliding-filament model of muscle contraction, based on myofibril experiments. Their work was published (*Nature* **173**, 973; 1954) alongside A. F. Huxley's *Nature* paper.

AGEING

Mixed results for dieting monkeys

According to previous studies, a low-calorie diet provides health benefits and increases lifespan in mammals, including primates. Yet a long-term investigation in rhesus monkeys finds no effect on longevity. [SEE LETTER P.318](#)

STEVEN N. AUSTAD

For more than 75 years, we have known that reducing food consumption increases longevity and preserves various aspects of health in mice and rats. For many researchers, these observations offered an investigational tool for probing underlying mechanisms of the ageing process. For other researchers and for the public at large, however, the observations suggested the more immediate hope that calorie restriction might improve health and extend life in humans — the fountain of youth could be as close as one's dinner plate.

To address whether such simple dietary intervention might have beneficial effects in primates that are long-lived like humans, two independent studies using rhesus monkeys were initiated in the late 1980s, one at the US National Institute on Aging (NIA) in Bethesda, Maryland, and the other at the Wisconsin National Primate Research Center (WNPRC) in Madison. Periodic reports from the studies have provided some indications of improved health^{1,2} and, possibly, increased lifespan³ in the animals. However, on page 318 of this issue, Mattison *et al.*⁴ report that the NIA data indicate that calorie restriction has no effect on the animals' longevity*.

In the NIA study, rhesus monkeys of different ages were subjected to a 30% calorie restriction (Fig. 1). The authors grouped the data into a young-onset group (animals that were 1–14 years old when enrolled in the study) and an old-onset group (aged 16–23 when enrolled). To give perspective to these ages, sexual maturity in captive rhesus monkeys occurs at 4–5 years of age, and median longevity had previously been reported⁵ as 27 years. Mattison *et al.* found that, for both age groups, the treated animals lived no longer than untreated monkeys. Nearly 50% of the young-onset animals are still alive, but the researchers' projections from current survival patterns indicate a likelihood of less than 0.1% that the calorie-restricted animals will survive longer than the controls.

To explore possible health benefits of the treatment, the authors regularly took blood



Figure 1 | Thinner but not longer lived. Mattison *et al.*⁴ report that restricting calorie intake in rhesus monkeys does not increase survival, although it improves certain metabolic parameters in some of the animals. The image shows two 27-year-old male monkeys, one of which had been placed on a calorie-restricted diet (left), with the other being untreated. (Housing for photo purposes only. Photo courtesy of the National Institute on Aging/NIH.)

samples from the animals to measure levels of glucose, cholesterol and triglyceride lipids. Cholesterol levels were significantly lower in treated males than in untreated males, but were unaffected in females. Moreover, the calorie-restricted old-onset monkeys had significantly lower triglyceride levels, as well as marginally lower glucose, than the controls. However, the young-onset treated animals showed no difference in glucose concentrations, and their triglyceride levels were marginally lower in males only, when compared with untreated controls of the same age group.

Calorie restriction did not alter causes of death. It may be premature to predict the final effect on ageing-related diseases because half of the animals in the young-onset group are still alive, but the dietary intervention seems to have reduced the incidence of cancer and, possibly, diabetes — although it might have slightly increased the incidence of cardiovascular disease. All of these ageing-related diseases did appear marginally later in the restricted monkeys, however.

Mattison and colleagues' results contrast

in some ways with those reported for the WNPRC study³, which enrolled only young rhesus monkeys (7–14 years old). The WNPRC researchers found no statistically significant effect of calorie restriction on overall longevity, but the survival patterns (with roughly 50% of the animals still alive) suggested a clear trend towards longer life for the treated subjects. Moreover, when deaths due to non-ageing-related causes such as gastrointestinal bloat or anaesthesia (which was required for some experimental procedures) were eliminated from the WNPRC analysis, calorie restriction did show a statistically significant, positive effect on survival.

However, some of us in the field were concerned about the interpretation of the WNPRC study. The deaths that had been eliminated from the analysis made up nearly half of the total deaths, and there were more of such deaths in the restricted group than in the control group. So, it was possible that some of the non-ageing-related deaths might be linked to the treatment. Interestingly, Mattison and colleagues' data also indicate a higher incidence of

*This article and the paper under discussion⁴ were published online on 29 August 2012.

non-ageing-related deaths in calorie-restricted young-onset animals than in controls.

What might account for the differences between the studies' results? One possibility is diet composition. Whereas the diets were broadly similar in their overall content of carbohydrates, proteins and fats, they differed in the specific types of such nutrients. For instance, sucrose made up 28.5% of the WNPRC diet, but only 3.9% of the NIA diet. Possibly related to this difference, more than 40% of the WNPRC control animals and only 12.5% of the NIA controls developed diabetes — although this metabolic malfunction was completely absent in the WNPRC calorie-restricted animals, but not in the treated animals in the NIA study.

Another difference is that the NIA controls were given an apportioned amount of food to prevent obesity, whereas the WNPRC controls were fed *ad libitum* (that is, they could eat as much as they pleased). Consequently, NIA control animals weighed less and were considerably longer-lived than the WNPRC controls. One interpretation of this observation is that the NIA controls were partially restricted, which would account for the lack of a survival effect of the treatment in the NIA study. Nevertheless, all animals in both studies — even in the calorie-restricted groups — weighed more than wild-caught monkeys⁶.

Taken together, the contrasting results raise an intriguing question about the nature and robustness of restricting calorie intake in primates. Is calorie restriction anything more than the elimination of excess fat? That might be concluded if one interprets the control animals in the NIA study as being restricted to a healthy weight, such that further restriction had little additional effect on longevity. However, researchers who use rodent models have long assumed that calorie restriction results in more than just leanness and that it extends life beyond its normal limits. It is known, for instance, that calorie restriction improves survival in obese and non-obese mice and rats⁷, even at very high restriction levels⁸, and that survival is affected by calorie restriction in a different way from exercise-driven leanness⁹.

The same question arises in ongoing studies comparing overweight^{10,11} or lean¹² people with normal-weight controls. If calorie restriction defines only the food intake needed to maintain a healthy body weight, then pharmacological mimicry of such an effect might improve the health of only the fraction of the population that is overweight. And, in that case — somewhat disappointingly — no spectacular increase in health or longevity should be expected. ■

Steven N. Austad is at the Barshop Institute for Longevity and Aging Studies, Department of Cellular and Structural Biology, University of Texas Health Science

Center at San Antonio, San Antonio, Texas 78245, USA.
e-mail: austad@uthscsa.edu

1. Messaoudi, I. *et al. Proc. Natl Acad. Sci. USA* **103**, 19448–19453 (2006).
2. Colman, R. J., Beasley, T. M., Allison, D. B. & Weindruch, R. *J. Gerontol. A* **63**, 556–559 (2008).
3. Colman, R. J. *et al. Science* **325**, 201–204 (2009).
4. Mattison, J. A. *et al. Nature* **489**, 318–321 (2012).
5. Colman, R. J. & Kemnitz, J. W. in *Methods in Aging Research* (ed. Yu, B. P.) 249–267 (CRC Press, 1998).

6. Leutenegger, W. & Cheverud, J. *Int. J. Primatol.* **3**, 387–402 (1982).
7. Turturro, A. *et al. J. Gerontol. A* **54**, B492–B501 (1999).
8. Weindruch, R., Walford, R. L., Fligiel, S. & Guthrie, D. *J. Nutr.* **116**, 641–654 (1986).
9. Holloszy, J. O. *J. Appl. Physiol.* **82**, 399–403 (1997).
10. Heilbronn, L. K. *et al. J. Am. Med. Assoc.* **295**, 1539–1548 (2006).
11. Fontana, L. *et al. Am. J. Physiol. Endocrinol. Metab.* **293**, E197–E202 (2007).
12. Fontana, L., Meyer, T. E., Klein, S. & Holloszy, J. O. *Proc. Natl Acad. Sci. USA* **101**, 6659–6663 (2004).

BIOGEOCHEMISTRY

Drought and tropical soil emissions

Past research implied that positive feedback might exist between climate change and greenhouse-gas emissions from soil. A study finds that drought-induced declines in such emissions from tropical forests could counter climate change.

CORY C. CLEVELAND
& BENJAMIN W. SULLIVAN

Tropical forests have been aptly described as the planet's lungs. Tropical vegetation inhales vast amounts of carbon dioxide and exhales oxygen to the atmosphere, and in so doing, provides a vital climate benefit by absorbing some of the nearly 9 petagrams (1 petagram is 10^{15} g) of carbon that human activities release into the atmosphere each year¹. What's more, tropical soils influence climate by releasing large amounts of CO₂ as a by-product of the decomposition of organic matter, and through several other soil microbial processes that both emit and consume greenhouse gases, including nitrous oxide and methane².

Important questions remain about the effect of climate change on greenhouse-gas fluxes from soil, and hence to what extent the beneficial climate services provided by tropical forests will persist in the future. Writing in *Global Biogeochemical Cycles*, Wood and Silver³ report that an experimentally induced drought had reduced greenhouse-gas emissions from tropical soil to the atmosphere, an effect that represents a negative feedback to climate change.

Many of the scenarios used to predict future climate focus on rising temperature, but changes in the water balance of tropical forests might have more immediate and pronounced effects. Although substantial uncertainty remains, most climate models predict significant declines in rainfall across large areas of the tropics^{4,5}. So how might such declines alter the net balance of the gases that contribute to climate change?

To address this question, Wood and Silver manipulated rainfall in a Puerto Rican tropical

forest (Fig. 1) to test how drought affects soil emissions of CO₂, methane and nitrous oxide. Such experimental manipulations remain the gold standard for directly assessing the potential effects of environmental change on ecosystem processes, and can provide critical input to the development of better Earth-system models. The authors simulated the effects of a prolonged dry season by preventing rainfall from reaching the forest floor for three months, and measured greenhouse-gas emissions from soil before, during and after the experiment.

After the simulated drought, Wood and Silver documented profound declines in the efflux of CO₂ from soil and increases in soil consumption of methane. In addition, the drought elicited surprising reductions in denitrification (the production of nitrous oxide), a process that is carried out by a tiny fraction of the soil's microbial community. Together, these effects lowered the global-warming potential of the soil emissions. When the heat-trapping ability of nitrous oxide is taken into account, a substantial reduction in total global-warming potential from all three gases was accounted for by the decline of denitrification. Because tropical forest soils are a major source of nitrous oxide emissions to the atmosphere⁶, this result implies that a drier climate could cause substantial reductions in global emissions of nitrous oxide from soil.

However, extrapolating results from experiments such as these is particularly challenging for tropical forests. In Wood and Silver's experiment, drought effects varied depending on landscape position: ridge and valley plots had larger responses to experimental drought than did those on slopes. This variation has considerable implications for models used to simulate biogeochemical responses to climate

non-ageing-related deaths in calorie-restricted young-onset animals than in controls.

What might account for the differences between the studies' results? One possibility is diet composition. Whereas the diets were broadly similar in their overall content of carbohydrates, proteins and fats, they differed in the specific types of such nutrients. For instance, sucrose made up 28.5% of the WNPRC diet, but only 3.9% of the NIA diet. Possibly related to this difference, more than 40% of the WNPRC control animals and only 12.5% of the NIA controls developed diabetes — although this metabolic malfunction was completely absent in the WNPRC calorie-restricted animals, but not in the treated animals in the NIA study.

Another difference is that the NIA controls were given an apportioned amount of food to prevent obesity, whereas the WNPRC controls were fed *ad libitum* (that is, they could eat as much as they pleased). Consequently, NIA control animals weighed less and were considerably longer-lived than the WNPRC controls. One interpretation of this observation is that the NIA controls were partially restricted, which would account for the lack of a survival effect of the treatment in the NIA study. Nevertheless, all animals in both studies — even in the calorie-restricted groups — weighed more than wild-caught monkeys⁶.

Taken together, the contrasting results raise an intriguing question about the nature and robustness of restricting calorie intake in primates. Is calorie restriction anything more than the elimination of excess fat? That might be concluded if one interprets the control animals in the NIA study as being restricted to a healthy weight, such that further restriction had little additional effect on longevity. However, researchers who use rodent models have long assumed that calorie restriction results in more than just leanness and that it extends life beyond its normal limits. It is known, for instance, that calorie restriction improves survival in obese and non-obese mice and rats⁷, even at very high restriction levels⁸, and that survival is affected by calorie restriction in a different way from exercise-driven leanness⁹.

The same question arises in ongoing studies comparing overweight^{10,11} or lean¹² people with normal-weight controls. If calorie restriction defines only the food intake needed to maintain a healthy body weight, then pharmacological mimicry of such an effect might improve the health of only the fraction of the population that is overweight. And, in that case — somewhat disappointingly — no spectacular increase in health or longevity should be expected. ■

Steven N. Austad is at the Barshop Institute for Longevity and Aging Studies, Department of Cellular and Structural Biology, University of Texas Health Science

Center at San Antonio, San Antonio, Texas 78245, USA.
e-mail: austad@uthscsa.edu

1. Messaoudi, I. *et al. Proc. Natl Acad. Sci. USA* **103**, 19448–19453 (2006).
2. Colman, R. J., Beasley, T. M., Allison, D. B. & Weindruch, R. *J. Gerontol. A* **63**, 556–559 (2008).
3. Colman, R. J. *et al. Science* **325**, 201–204 (2009).
4. Mattison, J. A. *et al. Nature* **489**, 318–321 (2012).
5. Colman, R. J. & Kemnitz, J. W. in *Methods in Aging Research* (ed. Yu, B. P.) 249–267 (CRC Press, 1998).

6. Leutenegger, W. & Cheverud, J. *Int. J. Primatol.* **3**, 387–402 (1982).
7. Turturro, A. *et al. J. Gerontol. A* **54**, B492–B501 (1999).
8. Weindruch, R., Walford, R. L., Fligiel, S. & Guthrie, D. *J. Nutr.* **116**, 641–654 (1986).
9. Holloszy, J. O. *J. Appl. Physiol.* **82**, 399–403 (1997).
10. Heilbronn, L. K. *et al. J. Am. Med. Assoc.* **295**, 1539–1548 (2006).
11. Fontana, L. *et al. Am. J. Physiol. Endocrinol. Metab.* **293**, E197–E202 (2007).
12. Fontana, L., Meyer, T. E., Klein, S. & Holloszy, J. O. *Proc. Natl Acad. Sci. USA* **101**, 6659–6663 (2004).

BIOGEOCHEMISTRY

Drought and tropical soil emissions

Past research implied that positive feedback might exist between climate change and greenhouse-gas emissions from soil. A study finds that drought-induced declines in such emissions from tropical forests could counter climate change.

CORY C. CLEVELAND
& BENJAMIN W. SULLIVAN

Tropical forests have been aptly described as the planet's lungs. Tropical vegetation inhales vast amounts of carbon dioxide and exhales oxygen to the atmosphere, and in so doing, provides a vital climate benefit by absorbing some of the nearly 9 petagrams (1 petagram is 10^{15} g) of carbon that human activities release into the atmosphere each year¹. What's more, tropical soils influence climate by releasing large amounts of CO₂ as a by-product of the decomposition of organic matter, and through several other soil microbial processes that both emit and consume greenhouse gases, including nitrous oxide and methane².

Important questions remain about the effect of climate change on greenhouse-gas fluxes from soil, and hence to what extent the beneficial climate services provided by tropical forests will persist in the future. Writing in *Global Biogeochemical Cycles*, Wood and Silver³ report that an experimentally induced drought had reduced greenhouse-gas emissions from tropical soil to the atmosphere, an effect that represents a negative feedback to climate change.

Many of the scenarios used to predict future climate focus on rising temperature, but changes in the water balance of tropical forests might have more immediate and pronounced effects. Although substantial uncertainty remains, most climate models predict significant declines in rainfall across large areas of the tropics^{4,5}. So how might such declines alter the net balance of the gases that contribute to climate change?

To address this question, Wood and Silver manipulated rainfall in a Puerto Rican tropical

forest (Fig. 1) to test how drought affects soil emissions of CO₂, methane and nitrous oxide. Such experimental manipulations remain the gold standard for directly assessing the potential effects of environmental change on ecosystem processes, and can provide critical input to the development of better Earth-system models. The authors simulated the effects of a prolonged dry season by preventing rainfall from reaching the forest floor for three months, and measured greenhouse-gas emissions from soil before, during and after the experiment.

After the simulated drought, Wood and Silver documented profound declines in the efflux of CO₂ from soil and increases in soil consumption of methane. In addition, the drought elicited surprising reductions in denitrification (the production of nitrous oxide), a process that is carried out by a tiny fraction of the soil's microbial community. Together, these effects lowered the global-warming potential of the soil emissions. When the heat-trapping ability of nitrous oxide is taken into account, a substantial reduction in total global-warming potential from all three gases was accounted for by the decline of denitrification. Because tropical forest soils are a major source of nitrous oxide emissions to the atmosphere⁶, this result implies that a drier climate could cause substantial reductions in global emissions of nitrous oxide from soil.

However, extrapolating results from experiments such as these is particularly challenging for tropical forests. In Wood and Silver's experiment, drought effects varied depending on landscape position: ridge and valley plots had larger responses to experimental drought than did those on slopes. This variation has considerable implications for models used to simulate biogeochemical responses to climate



Figure 1 | Luquillo Experimental Forest. Wood and Silver³ have studied the effect of drought on greenhouse gases emitted from soils in the Luquillo Experimental Forest, Puerto Rico.

change, none of which is able to capture the complex interplay of factors that can create substantial biogeochemical heterogeneity within and across tropical forests⁷. The authors' results are a prime example of the challenge facing scientists — many tropical regions have highly complex landscapes with varying nutrient availability, which in turn can regulate biological processes that influence greenhouse-gas production^{8,9}.

The possibility of positive feedback between climate change and greenhouse-gas emissions from soil has been recognized for decades. Wood and Silver's results suggest the opposite: declines in greenhouse-gas emissions following drought would reduce climate forcing. But as the authors highlight, the handful of studies in which rainfall has been excluded from tropical forests have shown positive, negative or no net effects of drought on soil greenhouse-gas emissions. Moreover, although it is crucial to consider the effects of drought alone, real-world emissions will hinge on the combined effects of changing precipitation and temperature, along with chronic shifts in atmospheric CO₂ levels and nutrient deposition — factors that were not manipulated in the authors' experiment. Finally, studies^{10,11} have shown that the growth and carbon uptake of trees in the tropics are highly sensitive to climate, but the experimental plots used in Wood and Silver's study (1.54 square metres) were not large enough to simulate the potentially negative effects of drought on carbon uptake through tree growth. The overall effects of drought on the greenhouse-gas balance of the sites therefore remain unknown.

In recent decades, tropical forests have given us a discount on anthropogenic CO₂ emissions by absorbing more greenhouse gases than human activity produces. Wood and Silver's findings suggest that tropical soils may continue to offset greenhouse-gas emissions during drought. Perhaps more notably, their study highlights the need for additional

large-scale experiments that can more completely resolve the potential effects of climate change on trace gas emissions in tropical forests. The authors' data are critical for improving and validating models that predict ecosystem and climate responses over large spatial and temporal scales. Considering the pivotal role of tropical ecosystems in basic human health and

welfare, our understanding of fundamental ecosystem processes, and their potential response to climate change, remains woefully incomplete in complex and diverse tropical forests. ■

Cory C. Cleveland and Benjamin W. Sullivan are in the Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, Montana 59812, USA.

e-mails: cory.cleveland@cfc.umt.edu; benjamin.sullivan@cfc.umt.edu

1. Le Quéré, C. *et al.* *Nature Geosci.* **2**, 831–836 (2009).
2. Montzka, S. A., Dlugokencky, E. J. & Butler, J. H. *Nature* **476**, 43–50 (2011).
3. Wood, T. E. & Silver, W. L. *Glob. Biogeochem. Cycles* **26**, GB3005 <http://dx.doi.org/10.1029/2010GB004014> (2012).
4. Cox, P. M. *et al.* *Theor. Appl. Climatol.* **78**, 137–156 (2004).
5. Neelin, J. D., Münnich, M., Su, H., Meyerson, J. E. & Holloway, C. E. *Proc. Natl Acad. Sci. USA* **103**, 6110–6115 (2006).
6. Matson, P. A. & Vitousek, P. M. *BioScience* **40**, 667–672 (1990).
7. Townsend, A. R., Asner, G. P. & Cleveland, C. C. *Trends Ecol. Evol.* **23**, 424–431 (2008).
8. Vitousek, P. M. *et al.* *Ecosystems* **6**, 762–772 (2003).
9. Porder, S. & Hilley, G. E. *Biogeochemistry* **102**, 153–166 (2011).
10. Clark, D. A., Piper, S. C., Keeling, C. D. & Clark, D. B. *Proc. Natl Acad. Sci. USA* **100**, 5852–5857 (2003).
11. Zhao, M. & Running, S. W. *Science* **329**, 940–943 (2010).

SOCIAL SCIENCE

Poked to vote

A Facebook message sent out during the 2010 US congressional elections influenced the voting behaviour of millions of people. The experiment illustrates the power of digital social networks to spread behavioural change. [SEE LETTER P.295](#)

SINAN ARAL

Social networks are the pathways through which information, advice, resources and support flow between people. They are essential for many people's decision-making, cooperation and complex interdependence. Yet although humans have almost always lived in networks, advances in computing power and new social technologies have only recently facilitated the development of forms of networked communication that are automating and accelerating the social signals that pulse through the human network on a daily basis. The rapid dissemination of social signals in these digital networks — status updates, tweets, likes, posts, shares and so on — raises serious scientific questions: how, when and to what extent do these signals influence decision-making and the spread of behaviours in society? If social influence drives behaviour, then digital social signals could be used to promote widespread behaviour change and thus to transform commerce, politics and

public health. On page 295 of this issue, Bond *et al.*¹ present some of the most convincing evidence to date that peer influence and digital social signals can affect political mobilization.

Political mobilization has been central to recent discourse about the transformative effects of social media — for example, the part that technologies such as Facebook or Twitter played in the protests collectively known as the Arab Spring, or may play in the forthcoming US presidential election. The question is: what role do peer influence and digital social signals have in mobilizing political expression? Do our friends' behaviours inspire us to be politically active, to protest or to vote?

These questions may seem relatively simple to answer, given the right data. But several statistical challenges make it difficult to quantitatively estimate peer influence in networks. For example, networks are homophilous — we tend to make friends with people like ourselves and thus have preferences that are highly correlated with those of our friends². If two friends adopt a behaviour, one immediately after the



Figure 1 | Luquillo Experimental Forest. Wood and Silver³ have studied the effect of drought on greenhouse gases emitted from soils in the Luquillo Experimental Forest, Puerto Rico.

change, none of which is able to capture the complex interplay of factors that can create substantial biogeochemical heterogeneity within and across tropical forests⁷. The authors' results are a prime example of the challenge facing scientists — many tropical regions have highly complex landscapes with varying nutrient availability, which in turn can regulate biological processes that influence greenhouse-gas production^{8,9}.

The possibility of positive feedback between climate change and greenhouse-gas emissions from soil has been recognized for decades. Wood and Silver's results suggest the opposite: declines in greenhouse-gas emissions following drought would reduce climate forcing. But as the authors highlight, the handful of studies in which rainfall has been excluded from tropical forests have shown positive, negative or no net effects of drought on soil greenhouse-gas emissions. Moreover, although it is crucial to consider the effects of drought alone, real-world emissions will hinge on the combined effects of changing precipitation and temperature, along with chronic shifts in atmospheric CO₂ levels and nutrient deposition — factors that were not manipulated in the authors' experiment. Finally, studies^{10,11} have shown that the growth and carbon uptake of trees in the tropics are highly sensitive to climate, but the experimental plots used in Wood and Silver's study (1.54 square metres) were not large enough to simulate the potentially negative effects of drought on carbon uptake through tree growth. The overall effects of drought on the greenhouse-gas balance of the sites therefore remain unknown.

In recent decades, tropical forests have given us a discount on anthropogenic CO₂ emissions by absorbing more greenhouse gases than human activity produces. Wood and Silver's findings suggest that tropical soils may continue to offset greenhouse-gas emissions during drought. Perhaps more notably, their study highlights the need for additional

large-scale experiments that can more completely resolve the potential effects of climate change on trace gas emissions in tropical forests. The authors' data are critical for improving and validating models that predict ecosystem and climate responses over large spatial and temporal scales. Considering the pivotal role of tropical ecosystems in basic human health and

welfare, our understanding of fundamental ecosystem processes, and their potential response to climate change, remains woefully incomplete in complex and diverse tropical forests. ■

Cory C. Cleveland and Benjamin W. Sullivan are in the Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, Montana 59812, USA.

e-mails: cory.cleveland@cfc.umt.edu; benjamin.sullivan@cfc.umt.edu

1. Le Quéré, C. *et al.* *Nature Geosci.* **2**, 831–836 (2009).
2. Montzka, S. A., Dlugokencky, E. J. & Butler, J. H. *Nature* **476**, 43–50 (2011).
3. Wood, T. E. & Silver, W. L. *Glob. Biogeochem. Cycles* **26**, GB3005 <http://dx.doi.org/10.1029/2010GB004014> (2012).
4. Cox, P. M. *et al.* *Theor. Appl. Climatol.* **78**, 137–156 (2004).
5. Neelin, J. D., Münnich, M., Su, H., Meyerson, J. E. & Holloway, C. E. *Proc. Natl Acad. Sci. USA* **103**, 6110–6115 (2006).
6. Matson, P. A. & Vitousek, P. M. *BioScience* **40**, 667–672 (1990).
7. Townsend, A. R., Asner, G. P. & Cleveland, C. C. *Trends Ecol. Evol.* **23**, 424–431 (2008).
8. Vitousek, P. M. *et al.* *Ecosystems* **6**, 762–772 (2003).
9. Porder, S. & Hilley, G. E. *Biogeochemistry* **102**, 153–166 (2011).
10. Clark, D. A., Piper, S. C., Keeling, C. D. & Clark, D. B. *Proc. Natl Acad. Sci. USA* **100**, 5852–5857 (2003).
11. Zhao, M. & Running, S. W. *Science* **329**, 940–943 (2010).

SOCIAL SCIENCE

Poked to vote

A Facebook message sent out during the 2010 US congressional elections influenced the voting behaviour of millions of people. The experiment illustrates the power of digital social networks to spread behavioural change. [SEE LETTER P.295](#)

SINAN ARAL

Social networks are the pathways through which information, advice, resources and support flow between people. They are essential for many people's decision-making, cooperation and complex interdependence. Yet although humans have almost always lived in networks, advances in computing power and new social technologies have only recently facilitated the development of forms of networked communication that are automating and accelerating the social signals that pulse through the human network on a daily basis. The rapid dissemination of social signals in these digital networks — status updates, tweets, likes, posts, shares and so on — raises serious scientific questions: how, when and to what extent do these signals influence decision-making and the spread of behaviours in society? If social influence drives behaviour, then digital social signals could be used to promote widespread behaviour change and thus to transform commerce, politics and

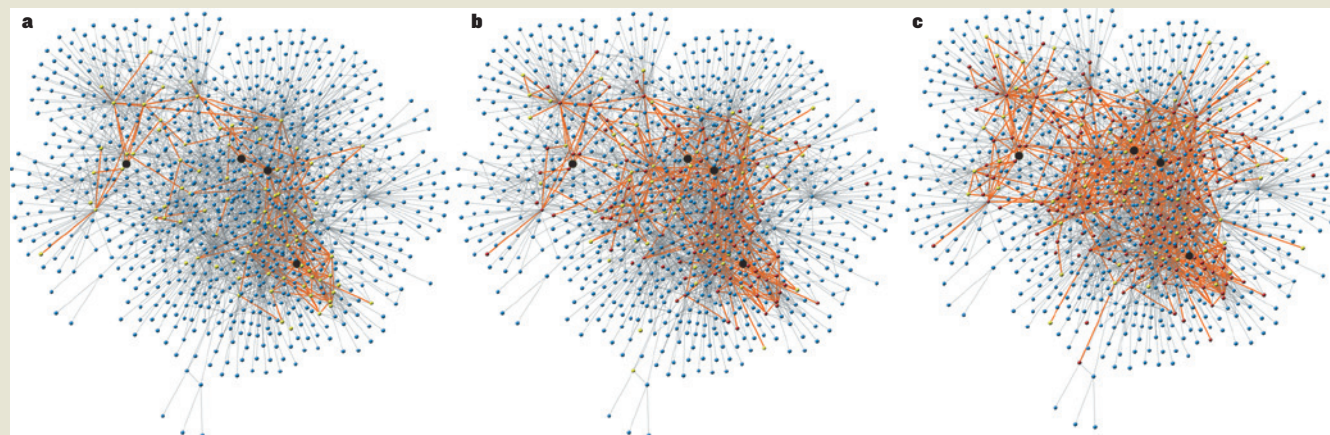
public health. On page 295 of this issue, Bond *et al.*¹ present some of the most convincing evidence to date that peer influence and digital social signals can affect political mobilization.

Political mobilization has been central to recent discourse about the transformative effects of social media — for example, the part that technologies such as Facebook or Twitter played in the protests collectively known as the Arab Spring, or may play in the forthcoming US presidential election. The question is: what role do peer influence and digital social signals have in mobilizing political expression? Do our friends' behaviours inspire us to be politically active, to protest or to vote?

These questions may seem relatively simple to answer, given the right data. But several statistical challenges make it difficult to quantitatively estimate peer influence in networks. For example, networks are homophilous — we tend to make friends with people like ourselves and thus have preferences that are highly correlated with those of our friends². If two friends adopt a behaviour, one immediately after the

BOX 1

Networking behaviour



The images shown display the uptake of a personalized mobile-phone application across a network of Yahoo! customers communicating over the company's instant-messenger service³. Each panel displays a snapshot of one cluster of the network at 33 days (a), 41 days (b) and 62 days (c) after the application's launch. The four large black nodes represent customers who were initial adopters of the application; yellow nodes are new adopters in each period; red

nodes are those who adopted in previous periods; and blue nodes are non-adopters. Orange lines highlight connections between previous adopters and new adopters that are activated as the product diffuses through the population.

A key question is how much the spread of behaviours such as product adoption is due to social influence — or homophily (correlated preferences among friends) — and/or other confounding factors.

Such confounding factors mimic the influence that the spread of behaviours in social networks has on the resulting diffusion patterns. Separating that influence from homophily and other factors, and establishing estimates of the degree to which peer-induced action is responsible for the adoption of a behaviour, are essential for creating effective policies that promote positive behaviours in society and contain negative ones. [S.A.](#)

other, it is difficult to distinguish whether this is due to peer influence or homophily. Network peers are also likely to be exposed to similar environmental stimuli: we work together and thus receive the same employee incentives; we live together and so are exposed to the same neighbourhood events; and because we have the same preferences, we tend to browse the same websites and watch the same television shows, so we see the same advertisements. This correlated exposure to external stimuli could also contribute to correlations in peer behaviours over time.

Methods are now available that can separate peer influence from homophily and other confounding factors in observational data³. This is useful because most data on these questions are observational. But controlling for unobservable factors, such as latent homophily, remains difficult⁴. As an alternative to observational data, experimental network studies using random assignment can provide a more robust means of identifying causal peer effects in networks and distinguishing between influence and confounding factors. Such experiments have demonstrated a role for peer influence in product adoption^{5–7} (Box 1), health behaviours⁸ and altruism⁹.

However, although it has been postulated that voting is also motivated by interactions with friends, neighbours and family members¹⁰, most research and many 'get out the

vote' mobilization campaigns have focused on the individual^{11,12} (for an exception, see ref. 13 for experimental evidence of peer influence on voting behaviour in two-person households). By contrast, Bond and colleagues used a randomized experiment — involving 61 million people within the intricate social network of Facebook — to show that digital social signals directly influence political self-expression, information-seeking and real-world voting behaviour.

The experiment was conducted during the 2010 US congressional elections. The targeted Facebook users saw a statement at the top of their 'News Feed' that encouraged them to vote, provided a link to information on local polling places, contained a clickable 'I voted' button and showed a counter of Facebook users who had voted (see Fig. 1 of the paper¹). This was the 'informational' message, received by a control group of users. Another group received a 'social' message, which additionally included six randomly selected profile pictures of Facebook friends of the user who had already clicked the 'I voted' button. The researchers categorized these friends into either ordinary friends or close friends on the basis of the degree of Facebook interaction.

The results show that users who received the social message were 2.08% more likely to report that they themselves voted, 0.26% more likely to seek information about a polling

location, and 0.39% more likely to actually vote than users receiving the informational message (the authors estimated actual voting using data on 6.3 million users in their sample that could be matched to publicly available voting records). The authors also examined 'contagion' effects of the messages on users who themselves received neither the social nor the informational message, but who had a friend who received a message. Here, in the validated voting sample, individuals were 0.224% more likely to vote, for each close friend who received a message, than they would have been had their friend received no message.

Although these estimates may seem small, they translate into significant numbers of votes. A social message saying that a Facebook friend had voted generated 886,000 additional 'expressed' votes (clicks on the 'I voted' button), and messages involving a close friend generated an additional 559,000 expressed votes. In terms of real-world voting, the authors were able to validate 282,000 additional votes cast by people receiving a message that a close Facebook friend had voted.

This work has important implications for our understanding of how signals in social networks influence an individual's behaviour. But there is more work to be done. In particular, future work modelling the degree to which contamination or leakage in networked experiments⁴ (also called interference¹⁴) affects

inference, examination of how peer influence varies across behaviours⁵, and evaluation of the social and structural conditions under which influence is more or less likely to propagate will all be essential to our understanding of the spread of behaviour change through human populations.

Advancing our understanding of peer influence in networks is the first step towards designing 'network interventions' that can promote positive behaviours in human populations, or contain negative ones¹⁵. It is perhaps obvious that this is relevant to, for example, targeted advertising and viral marketing. But such interventions also have the potential to promote positive social changes, such as

increasing the rate of HIV testing, reducing violence, improving adherence to exercise, or increasing political mobilization and awareness. In this way, the science of social influence may have dramatic implications for products, politics and public health. ■

Sinan Aral is at the Leonard N. Stern School of Business, New York University, New York, New York 10012, USA.
e-mail: sinan@stern.nyu.edu

1. Bond, R. M. *et al.* *Nature* **489**, 295–298 (2012).
2. McPherson, M., Smith-Lovin, L. & Cook, J. M. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
3. Aral, S., Muchnik, L. & Sundararajan, A. *Proc. Natl Acad. Sci. USA* **106**, 21544–21549 (2009).

4. Shalizi, C. R. & Thomas, A. C. *Sociol. Methods Res.* **40**, 211–239 (2011).
5. Aral, S. & Walker, D. *Mgmt Sci.* **57**, 1623–1639 (2011).
6. Aral, S. & Walker, D. *Science* **337**, 337–341 (2012).
7. Bakshy, E., Eckles, D., Yan, R. & Rosenn, I. *Proc. 13th ACM Conf. Electron. Comm.* 146–161 (ACM, 2012).
8. Centola, D. *Science* **329**, 1194–1197 (2010).
9. Leider, S., Möbius, M. M., Rosenblatt, T. & Do, Q.-A. *Q. J. Econ.* **124**, 1815–1851 (2009).
10. McClurg, S. D. *Am. Polit. Res.* **32**, 406–443 (2004).
11. Rosenstone, S. J. & Hansen, J. M. *Mobilization, Participation and Democracy in America* (Macmillan, 1993).
12. Green, D. P., Gerber, A. S. & Nickerson, D. W. *J. Polit.* **65**, 1083–1096 (2003).
13. Nickerson, D. W. *Am. Polit. Sci. Rev.* **102**, 49–57 (2008).
14. Aronow, P. M. & Samii, C. *Estimating Average Causal Effects Under General Interference* New York Univ. Work. Pap. (2012).
15. Valente, T. W. *Science* **337**, 49–53 (2012).

ORGANIC SYNTHESIS

A biochemical messenger made easily

Biochemicals known as prostaglandins are challenging targets for synthetic organic chemistry. Yet by channelling the reactivity of a simple reactant, a powerful synthesis of one such compound has been achieved. [SEE LETTER P.278](#)

ERIK J. SORENSEN

Chemical messengers called prostaglandins are present in nearly all mammalian tissues. These elusive molecules mediate an extraordinary number of biological processes — including the regulation of body temperature, the contraction and relaxation of the human uterus, the aggregation of platelets in blood and cellular responses to inflammation. They have therefore been the targets of wide-ranging research¹ since the 1930s.

In particular, their unique molecular architectures and great therapeutic potential have fired the creative imagination of synthetic organic chemists^{2,3}. Writing on page 278 of this issue, Coulthard *et al.*⁴ report one of the cleverest syntheses of one such molecule, prostaglandin F_{2α} (PGF_{2α}), to date. The conciseness of their approach may open up new opportunities for drug discovery.

Prostaglandins contain two fat-soluble hydrocarbon chains on opposite sides of a ring of five carbon atoms (Fig. 1). Nature synthesizes

PGF_{2α} and its analogues from arachidonic acid (a polyunsaturated fatty acid) and two molecules of oxygen, with a little help from some key enzymes. Although nearly every nucleated cell is capable of biosynthesizing prostaglandins, these compounds are short-lived and exceedingly difficult to isolate from biological samples. Since the late 1960s, chemists have therefore devised creative strategies for producing prostaglandins from simple chemicals in the laboratory. The development of dependable laboratory syntheses has contributed substantially to our knowledge of the compounds' remarkable range of pharmacological properties, as well as their relevance to human health.

Ideas about the preparation of PGF_{2α} have resulted in a diversity of pathways^{5–9} — a diversity that demonstrates how well the principles of organic chemistry can be used to plan and execute synthetic routes to structurally complex molecules (Fig. 1). The approach now reported by Coulthard *et al.* will be praised for its remarkable brevity and for the bold strategy that guided it. The structural relationship between the simple starting material, succinaldehyde, and the coveted target is distant, and yet the authors perceived that a union of two molecules of succinaldehyde in a single laboratory operation might produce an intermediate hemiacetal compound that is tantalizingly close to the structure of PGF_{2α} (Fig. 2).

The risky aspect of the authors' strategy concerns the intrinsic reactivity of succinaldehyde and its potential to take part in undesired, polymer-forming reactions. Few chemists, even those of adventurous spirit, would have believed that a useful synthesis of the hemiacetal could be achieved as Coulthard *et al.* anticipated, through the direct pairing of two molecules of succinaldehyde in sequential carbon–carbon–bond-forming reactions known as aldol reactions. The great thing is that, not only did this direct approach work well, but it also achieved high stereoselectivity — it produced mostly one mirror-image isomer (enantiomer) of the product. This is crucial, because the biological activity of prostaglandins depends on their enantiomeric form.

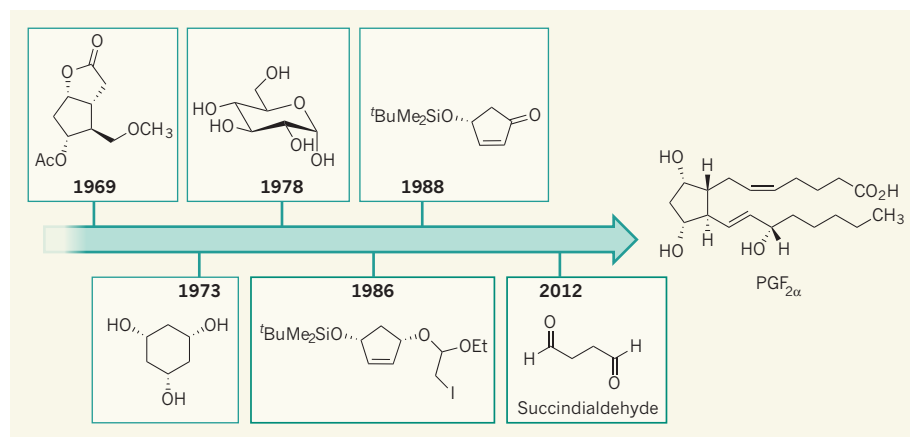


Figure 1 | Landmark syntheses of prostaglandin F_{2α} (PGF_{2α}). The PGF_{2α} biochemical has long acted as a testing ground for organic synthesis. The dates of selected syntheses^{5–9} are indicated, along with the molecule that acted as the starting point or key intermediate for each synthesis. Coulthard *et al.*⁴ now report a concise synthesis of PGF_{2α} that starts from succinaldehyde. Ac is acetyl, COCH₃; Me is methyl, CH₃; Et is ethyl, C₂H₅; 'Bu is tertiary butyl, (CH₃)₃C.

inference, examination of how peer influence varies across behaviours⁵, and evaluation of the social and structural conditions under which influence is more or less likely to propagate will all be essential to our understanding of the spread of behaviour change through human populations.

Advancing our understanding of peer influence in networks is the first step towards designing 'network interventions' that can promote positive behaviours in human populations, or contain negative ones¹⁵. It is perhaps obvious that this is relevant to, for example, targeted advertising and viral marketing. But such interventions also have the potential to promote positive social changes, such as

increasing the rate of HIV testing, reducing violence, improving adherence to exercise, or increasing political mobilization and awareness. In this way, the science of social influence may have dramatic implications for products, politics and public health. ■

Sinan Aral is at the Leonard N. Stern School of Business, New York University, New York, New York 10012, USA.
e-mail: sinan@stern.nyu.edu

1. Bond, R. M. *et al.* *Nature* **489**, 295–298 (2012).
2. McPherson, M., Smith-Lovin, L. & Cook, J. M. *Annu. Rev. Sociol.* **27**, 415–444 (2001).
3. Aral, S., Muchnik, L. & Sundararajan, A. *Proc. Natl Acad. Sci. USA* **106**, 21544–21549 (2009).

4. Shalizi, C. R. & Thomas, A. C. *Sociol. Methods Res.* **40**, 211–239 (2011).
5. Aral, S. & Walker, D. *Mgmt Sci.* **57**, 1623–1639 (2011).
6. Aral, S. & Walker, D. *Science* **337**, 337–341 (2012).
7. Bakshy, E., Eckles, D., Yan, R. & Rosenn, I. *Proc. 13th ACM Conf. Electron. Comm.* 146–161 (ACM, 2012).
8. Centola, D. *Science* **329**, 1194–1197 (2010).
9. Leider, S., Möbius, M. M., Rosenblatt, T. & Do, Q.-A. *Q. J. Econ.* **124**, 1815–1851 (2009).
10. McClurg, S. D. *Am. Polit. Res.* **32**, 406–443 (2004).
11. Rosenstone, S. J. & Hansen, J. M. *Mobilization, Participation and Democracy in America* (Macmillan, 1993).
12. Green, D. P., Gerber, A. S. & Nickerson, D. W. *J. Polit.* **65**, 1083–1096 (2003).
13. Nickerson, D. W. *Am. Polit. Sci. Rev.* **102**, 49–57 (2008).
14. Aronow, P. M. & Samii, C. *Estimating Average Causal Effects Under General Interference* New York Univ. Work. Pap. (2012).
15. Valente, T. W. *Science* **337**, 49–53 (2012).

ORGANIC SYNTHESIS

A biochemical messenger made easily

Biochemicals known as prostaglandins are challenging targets for synthetic organic chemistry. Yet by channelling the reactivity of a simple reactant, a powerful synthesis of one such compound has been achieved. [SEE LETTER P.278](#)

ERIK J. SORENSEN

Chemical messengers called prostaglandins are present in nearly all mammalian tissues. These elusive molecules mediate an extraordinary number of biological processes — including the regulation of body temperature, the contraction and relaxation of the human uterus, the aggregation of platelets in blood and cellular responses to inflammation. They have therefore been the targets of wide-ranging research¹ since the 1930s.

In particular, their unique molecular architectures and great therapeutic potential have fired the creative imagination of synthetic organic chemists^{2,3}. Writing on page 278 of this issue, Coulthard *et al.*⁴ report one of the cleverest syntheses of one such molecule, prostaglandin F_{2α} (PGF_{2α}), to date. The conciseness of their approach may open up new opportunities for drug discovery.

Prostaglandins contain two fat-soluble hydrocarbon chains on opposite sides of a ring of five carbon atoms (Fig. 1). Nature synthesizes

PGF_{2α} and its analogues from arachidonic acid (a polyunsaturated fatty acid) and two molecules of oxygen, with a little help from some key enzymes. Although nearly every nucleated cell is capable of biosynthesizing prostaglandins, these compounds are short-lived and exceedingly difficult to isolate from biological samples. Since the late 1960s, chemists have therefore devised creative strategies for producing prostaglandins from simple chemicals in the laboratory. The development of dependable laboratory syntheses has contributed substantially to our knowledge of the compounds' remarkable range of pharmacological properties, as well as their relevance to human health.

Ideas about the preparation of PGF_{2α} have resulted in a diversity of pathways^{5–9} — a diversity that demonstrates how well the principles of organic chemistry can be used to plan and execute synthetic routes to structurally complex molecules (Fig. 1). The approach now reported by Coulthard *et al.* will be praised for its remarkable brevity and for the bold strategy that guided it. The structural relationship between the simple starting material, succinaldehyde, and the coveted target is distant, and yet the authors perceived that a union of two molecules of succinaldehyde in a single laboratory operation might produce an intermediate hemiacetal compound that is tantalizingly close to the structure of PGF_{2α} (Fig. 2).

The risky aspect of the authors' strategy concerns the intrinsic reactivity of succinaldehyde and its potential to take part in undesired, polymer-forming reactions. Few chemists, even those of adventurous spirit, would have believed that a useful synthesis of the hemiacetal could be achieved as Coulthard *et al.* anticipated, through the direct pairing of two molecules of succinaldehyde in sequential carbon–carbon bond-forming reactions known as aldol reactions. The great thing is that, not only did this direct approach work well, but it also achieved high stereoselectivity — it produced mostly one mirror-image isomer (enantiomer) of the product. This is crucial, because the biological activity of prostaglandins depends on their enantiomeric form.

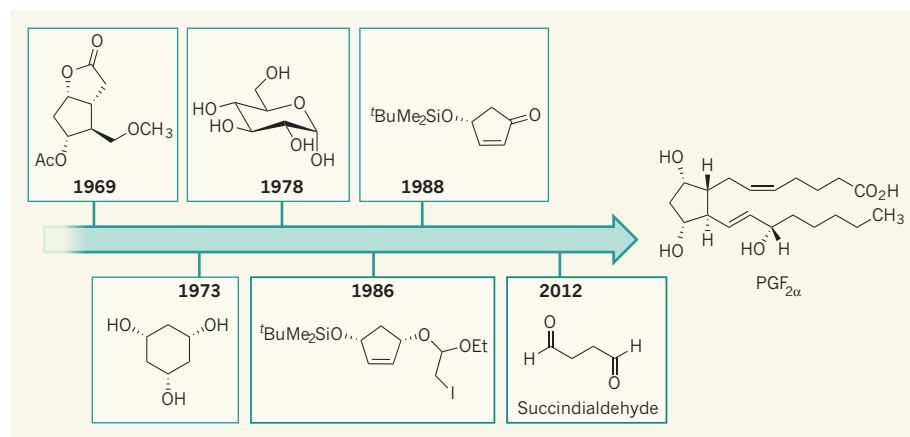


Figure 1 | Landmark syntheses of prostaglandin F_{2α} (PGF_{2α}). The PGF_{2α} biochemical has long acted as a testing ground for organic synthesis. The dates of selected syntheses^{5–9} are indicated, along with the molecule that acted as the starting point or key intermediate for each synthesis. Coulthard *et al.*⁴ now report a concise synthesis of PGF_{2α} that starts from succinaldehyde. Ac is acetyl, COCH₃; Me is methyl, CH₃; Et is ethyl, C₂H₅; 'Bu is tertiary butyl, (CH₃)₃C.

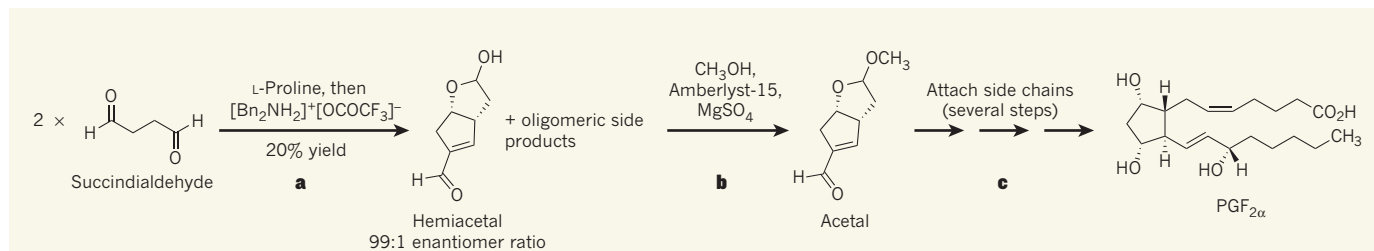


Figure 2 | A concise synthesis of PGF_{2α}. **a**, In Coulthard and colleagues' synthesis⁴, two equivalents of succinaldehyde react with each other to produce a hemiacetal when treated with catalytic L-proline, and then with another catalyst (an amine salt, [Bn₂NH₂]⁺[OCOCF₃]⁻), in the same reaction vessel. The product is obtained in 20% yield as a 99:1 ratio of enantiomers (mirror-image isomers); the major isomer is depicted.

Oligomeric side products are also formed, but are easily removed by filtration. Bn is benzyl, CH₂C₆H₅. **b**, The hemiacetal is treated with methanol (CH₃OH), an acid (Amberlyst-15) and magnesium sulphate (MgSO₄) to yield an acetal, which contains most of the features needed for the ring component of PGF_{2α}. **c**, The authors completed the synthesis by attaching two side chains to the acetal.

The authors discovered that the sequential addition of two organic catalysts to a solution of succinaldehyde is key to the formation of the desired hemiacetal. The first catalyst was the amino acid proline, and the second was an amine salt. The development of this process was supported by earlier studies of proline-catalysed aldol reactions^{10–12} and ring-forming aldol reactions induced by amine salts¹³ at room temperature.

Although the yield of hemiacetal isolated from the reaction was moderate (20%), Coulthard *et al.* removed most of the undesired oligomeric side products by filtration. A simple chromatographic purification step then afforded the product as the major component of a 99:1 mixture of enantiomers. Any possible concerns about the yield for this impressive transformation are assuaged by the authors' demonstration that 15 grams of a key prostaglandin building block (an acetal; Fig. 2) can be produced in only two steps from succinaldehyde — a large quantity to take forward for the rest of the synthesis.

With its ring of five carbon atoms and newly formed stereogenic centres (carbon atoms around which different geometric arrangements of bonds generate different enantiomers), the acetal building block is a close approximation of the core structure of the prostaglandins. By exploiting the reactive alkene group and the rigid molecular shape of the building block, Coulthard *et al.* attached one of the two hydrocarbon side chains of PGF_{2α} to the desired site on the five-membered ring. Then, in the final phase of the synthesis, the authors efficiently connected the second side chain using a reaction developed by E. J. Corey in his influential syntheses of prostaglandins⁵. In all, only seven operations were needed to transmute the simple compound succinaldehyde into PGF_{2α}. For comparison, Corey's synthesis required 17 steps, starting from the simple molecule cyclopentadiene.

NATURE.COM
For more on
prostaglandin
synthesis, see:
go.nature.com/tbqz85

Finding a way to prepare prostaglandins and their analogues in a few steps from commonplace chemicals is still highly

alluring for many researchers, especially those in the pharmaceutical industry. The brevity of Coulthard and colleagues' synthesis of PGF_{2α} sets a new standard, and their approach should be applicable to other members of the prostaglandin class, as well as to new structures that have some of the molecular features of prostaglandins. The authors' achievement also calls to mind a much earlier groundbreaking use of succinaldehyde in chemical synthesis — Robert Robinson's synthesis of tropinone¹⁴, a precursor to the drug atropine, in 1917. Some of the power and simplicity of that classic work is seen in Coulthard and colleagues' remarkable solution to the problem of building prostaglandins. ■

Erik J. Sorensen is in the Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.
e-mail: ej@s@princeton.edu

CELL BIOLOGY

Dormant and restless skin stem cells

It has been unclear whether a uniform group of stem cells gives rise to most cells in the epidermis. A study reveals the presence of at least two stem-cell populations that have different proliferative abilities. SEE ARTICLE P.257

LAURA DE ROSA & MICHELE DE LUCA

All renewing tissues, such as blood and skin, are sustained and repaired by a small population of resident stem cells. These cells have the ability to self-renew and to generate committed progenitor cells that differentiate into the cell lineages of the tissue of origin. However, the nature and the specific activities of stem and progenitor cells at different body sites are a matter of debate. For example, several theories have been proposed

regarding the epidermis, the outer covering of the skin (Fig. 1). On page 257 of this issue, Mascré *et al.*¹ provide compelling evidence for the existence of two cell populations that differ in their proliferative dynamics, their gene-expression profile and their ability to repair the epidermis after injury.

To investigate the origin, location, proliferation and fate of stem cells in mouse-tail epidermis, the authors used genetic lineage tracing — a technique that allows *in vivo* fluorescent marking of specific cell types and

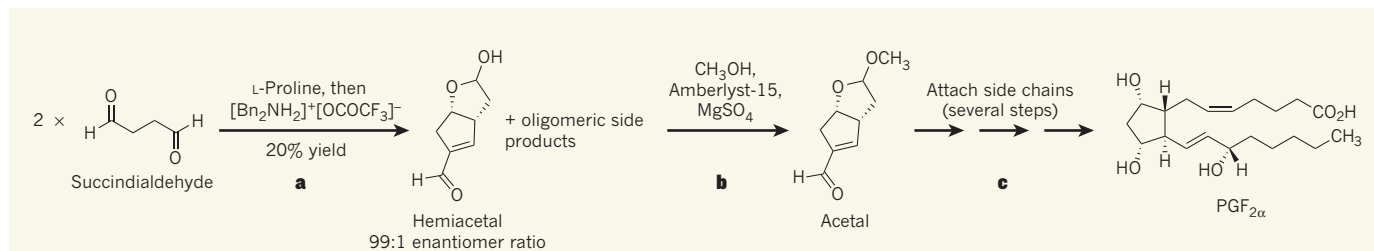


Figure 2 | A concise synthesis of PGF_{2α}. **a**, In Coulthard and colleagues' synthesis⁴, two equivalents of succinaldehyde react with each other to produce a hemiacetal when treated with catalytic L-proline, and then with another catalyst (an amine salt, [Bn₂NH₂]⁺[OCOCF₃]⁻), in the same reaction vessel. The product is obtained in 20% yield as a 99:1 ratio of enantiomers (mirror-image isomers); the major isomer is depicted.

Oligomeric side products are also formed, but are easily removed by filtration. Bn is benzyl, CH₂C₆H₅. **b**, The hemiacetal is treated with methanol (CH₃OH), an acid (Amberlyst-15) and magnesium sulphate (MgSO₄) to yield an acetal, which contains most of the features needed for the ring component of PGF_{2α}. **c**, The authors completed the synthesis by attaching two side chains to the acetal.

The authors discovered that the sequential addition of two organic catalysts to a solution of succinaldehyde is key to the formation of the desired hemiacetal. The first catalyst was the amino acid proline, and the second was an amine salt. The development of this process was supported by earlier studies of proline-catalysed aldol reactions^{10–12} and ring-forming aldol reactions induced by amine salts¹³ at room temperature.

Although the yield of hemiacetal isolated from the reaction was moderate (20%), Coulthard *et al.* removed most of the undesired oligomeric side products by filtration. A simple chromatographic purification step then afforded the product as the major component of a 99:1 mixture of enantiomers. Any possible concerns about the yield for this impressive transformation are assuaged by the authors' demonstration that 15 grams of a key prostaglandin building block (an acetal; Fig. 2) can be produced in only two steps from succinaldehyde — a large quantity to take forward for the rest of the synthesis.

With its ring of five carbon atoms and newly formed stereogenic centres (carbon atoms around which different geometric arrangements of bonds generate different enantiomers), the acetal building block is a close approximation of the core structure of the prostaglandins. By exploiting the reactive alkene group and the rigid molecular shape of the building block, Coulthard *et al.* attached one of the two hydrocarbon side chains of PGF_{2α} to the desired site on the five-membered ring. Then, in the final phase of the synthesis, the authors efficiently connected the second side chain using a reaction developed by E. J. Corey in his influential syntheses of prostaglandins⁵. In all, only seven operations were needed to transmute the simple compound succinaldehyde into PGF_{2α}. For comparison, Corey's synthesis required 17 steps, starting from the simple molecule cyclopentadiene.

NATURE.COM
For more on
prostaglandin
synthesis, see:
go.nature.com/tbqz85

Finding a way to prepare prostaglandins and their analogues in a few steps from commonplace chemicals is still highly

alluring for many researchers, especially those in the pharmaceutical industry. The brevity of Coulthard and colleagues' synthesis of PGF_{2α} sets a new standard, and their approach should be applicable to other members of the prostaglandin class, as well as to new structures that have some of the molecular features of prostaglandins. The authors' achievement also calls to mind a much earlier groundbreaking use of succinaldehyde in chemical synthesis — Robert Robinson's synthesis of tropinone¹⁴, a precursor to the drug atropine, in 1917. Some of the power and simplicity of that classic work is seen in Coulthard and colleagues' remarkable solution to the problem of building prostaglandins. ■

Erik J. Sorensen is in the Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA.
e-mail: ej@s@princeton.edu

CELL BIOLOGY

Dormant and restless skin stem cells

It has been unclear whether a uniform group of stem cells gives rise to most cells in the epidermis. A study reveals the presence of at least two stem-cell populations that have different proliferative abilities. SEE ARTICLE P.257

LAURA DE ROSA & MICHELE DE LUCA

All renewing tissues, such as blood and skin, are sustained and repaired by a small population of resident stem cells. These cells have the ability to self-renew and to generate committed progenitor cells that differentiate into the cell lineages of the tissue of origin. However, the nature and the specific activities of stem and progenitor cells at different body sites are a matter of debate. For example, several theories have been proposed

regarding the epidermis, the outer covering of the skin (Fig. 1). On page 257 of this issue, Mascré *et al.*¹ provide compelling evidence for the existence of two cell populations that differ in their proliferative dynamics, their gene-expression profile and their ability to repair the epidermis after injury.

To investigate the origin, location, proliferation and fate of stem cells in mouse-tail epidermis, the authors used genetic lineage tracing — a technique that allows *in vivo* fluorescent marking of specific cell types and

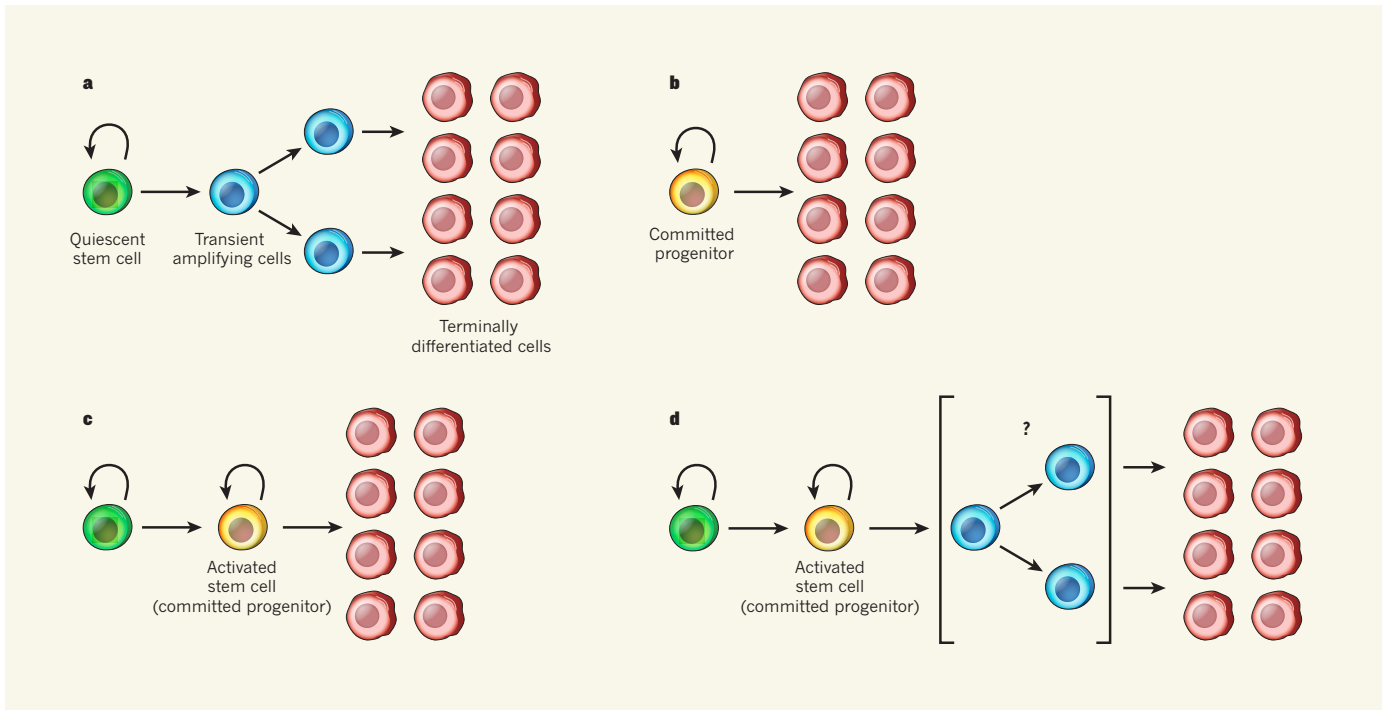


Figure 1 | Models for self-renewal of the epidermis. **a**, One model² of epidermal renewal relies on quiescent stem cells that generate transient actively dividing (amplifying) cells, which in turn generate non-dividing, differentiated cells. Curly arrows indicate self-renewing cells. **b**, A second model^{3,4} posits the existence of a single population of actively cycling, committed progenitor cells that directly generate differentiated cells. **c**, Mascré and colleagues' results¹ suggest that epidermal renewal is sustained by slowly dividing stem cells that persist during the lifetime

of the animal and contribute to both repair and long-term regeneration of the epidermis. These cells give rise to actively cycling, self-renewing stem cells (or committed progenitors) that persist for various periods of time, replenish differentiated cells and make only a short-term contribution to wound healing. **d**, We suggest that slowly dividing stem cells generate activated stem cells that give rise to transient amplifying progenitors, which eventually produce differentiated cells. A similar model has previously been proposed for other tissues^{5–10}.

their progeny — as well as elegant studies of individual clones (cells derived from a single initial cell). The specificity of the marking depends on the nature of the promoters (regulatory DNA sequences) that are employed to trigger permanent expression of a fluorescent protein. This is useful because different promoters are active in different cell types.

Mascré *et al.* used two distinct promoters (K14 and *Inv*) and unambiguously identified two cell populations. One of these populations consisted of slowly dividing cells, which represented the stem-cell reservoir and contributed to wound repair. These stem cells gave rise to a second population that comprised actively dividing committed progenitors, which eventually generated differentiated cells. The K14 promoter targeted both stem cells and committed progenitors, whereas the *Inv* promoter targeted only progenitors.

The results confirm the notion that the epidermis contains long-lived quiescent stem cells as well as cells endowed with differing proliferative abilities², and argues against the model proposing that a single progenitor-cell pool sustains self-renewal and regeneration of the mouse epidermis^{3,4} (Fig. 1a–c). Moreover, the authors' study provides an explanation for previous findings that seemed to suggest a single-progenitor model^{3,4}. These earlier studies had used a promoter (*Ach*) that targets just

one group of actively cycling cells that make only a short-term contribution to epidermal repair — in other words, the *Ach*-marked cells behaved in a similar way to Mascré and colleagues' *Inv*-targeted progenitor cells.

Interestingly, the two cell populations described by the authors both possess a fundamental stem-cell property, namely, self-renewal. However, the slowly cycling stem cells persist throughout the animal's lifetime and contribute to the repair and long-term regeneration of the epidermis, whereas the actively dividing progenitors endure for shorter periods (up to several months) and make only a short-term contribution to wound healing. Therefore, these actively dividing progenitors are not the canonical 'transient amplifying cells' proposed by a previous model (Fig. 1a).

How can we define the K14- and *Inv*-labelled cells from a functional point of view? To try to answer this question, it may be useful to discuss what happens in other tissues. It has been shown⁵ that the bone marrow contains a pool of quiescent haematopoietic (blood-forming) stem cells that generate a population of rapidly cycling stem cells. K14-marked stem cells are functionally similar to quiescent haematopoietic stem cells, whereas *Inv*-marked progenitors mirror actively cycling, self-renewing haematopoietic stem cells. A similar situation has been described

for tissues such as hair follicles and the small intestine^{6,7}. In all of these tissues, stem cells can switch between quiescence and activation depending on the functional status of the tissue; that is, normal self-renewal or acute wound healing. In our opinion, Mascré and colleagues' data support a model envisaging the coexistence of quiescent and activated stem cells as a general strategy for tissue renewal (Fig. 1c).

The notion that the mammalian epidermis contains cells endowed with different proliferative dynamics raises interesting questions that need further investigation. What factors specify the fate of these two cell populations? What are the molecular pathways that regulate stem-cell quiescence and activation in the epidermis, especially upon injury? And is the switch from quiescence to activation a cell-autonomous process, or is it controlled by environmental cues?

The study also highlights the limitations of genetic lineage tracing. Results obtained with this technique are influenced by the dynamics of the stem-cell populations in the tissues analysed, as well as by the types of promoter used and by how the promoters are activated. For instance, the authors did not detect transient amplifying progenitors, but it is not clear whether such cells are genuinely absent from the mouse epidermis or whether they were not detected owing to the lack

of an appropriate promoter (Fig. 1d).

Mascre and colleagues' work sheds light on the heterogeneity of the proliferative cell populations in mouse epidermis, and has increased our understanding of stem-cell biology. But does this knowledge apply to human tissues? Caution is required when inferring aspects of human tissue physiology from animal data. For instance, clonal analyses of several human squamous epithelia (tissues such as the epidermis, cornea and conjunctiva) have unambiguously shown the existence of self-renewing cells endowed with stem-cell properties, as well as non-self-renewing cells with differing capacities for multiplication, including canonical transient amplifying cells^{8,9}. Both cell types participate in the regeneration of these epithelia in the clinic¹⁰. ■

Laura De Rosa and Michele De Luca are at the Center for Regenerative Medicine 'Stefano Ferrari', University of Modena and Reggio Emilia, Modena 41125, Italy.
e-mail: michele.deluca@unimore.it

1. Mascre, G. *et al.* *Nature* **489**, 257–262 (2012).
2. Potten, C. S. *Cell Tissue Kinet.* **7**, 77–88 (1974).
3. Clayton, E. *et al.* *Nature* **446**, 185–189 (2007).
4. Doupe, D. P. *et al.* *Dev. Cell* **18**, 317–323 (2010).
5. Wilson, A. *et al.* *Cell* **135**, 1118–1129 (2008).
6. Barker, N. *et al.* *Nature* **449**, 1003–1007 (2007).
7. Jaks, V. *et al.* *Nature Genet.* **40**, 1291–1299 (2008).
8. Barrandon, Y. & Green, H. *Proc. Natl Acad. Sci. USA* **84**, 2302–2306 (1987).
9. Pellegrini, G. *et al.* *J. Cell Biol.* **145**, 769–782 (1999).
10. Rama, P. *et al.* *N. Engl. J. Med.* **363**, 147–155 (2010).

ENVIRONMENTAL SCIENCE

The rainforest's water pump

An investigation of naturally occurring water recycling in rainforests finally marries the results of global climate models with observations. Alarming, it also suggests that deforestation can greatly reduce tropical rainfall. [SEE LETTER P.282](#)

LUIZ E. O. C. ARAGÃO

The humid tropics contain more than 35% of global forests, covering an area of 11,564,000 square kilometres (ref. 1). Tropical trees can extract deep soil water and pump it back to the atmosphere through a process called evapotranspiration. As a result, 25–56% of the rainfall in these regions can be recycled within the ecosystem². Large-scale climate models indicate that this mechanism — which maintains atmospheric moisture and so feeds far-inland regions with rainfall — may be weakened by the removal of existing tropical forests^{3,4}. Yet, paradoxically, an increase in local rainfall over deforested areas has been observed⁵.

On page 282 of this issue, Spracklen *et al.*⁶ report an analysis of tropical rainfall that combines models of atmospheric transport with satellite observations of rainfall and vegetation cover*. The authors conclude that, for more than 60% of pan-tropical land (that is, the tropics across all continents), air masses that have travelled over extensive vegetated surfaces can generate at least twice as much rainfall as air masses that have flowed over deforested lands. On the basis of their findings, the authors predict a potentially widespread reduction of rainfall in the Amazon basin, if deforestation

in the region were to continue at the rate⁷ that occurred from 1997 to 2002.

Sustaining high rates of global human-population growth requires an increase in farmable land, energy and timber supplies. Without adequate planning, these needs tend to be met by cutting down rainforests. Because ecological processes and human activities (such as agriculture) depend on water provision, scientists have, for at least two decades, investigated the impact of tropical deforestation on the hydrological cycle, especially in Amazonia.

Pioneering studies^{3,4} using global climate models (GCMs) found that vegetation removal can disrupt the water cycle. For example, models show that changes in land surface characteristics cause a reduction in evapotranspiration rates and so induce basin-wide decreases in Amazonian rainfall³. Another simulation shows that removing around 40% of the original forest cover could drive Amazonia into an irreversible, drier climate mode⁸.

These findings, however, are at odds with observations. Studies indicate that rainfall over deforested areas is, in fact, higher than in adjacent forests⁵. This is because increased surface heating in deforested areas induces upward air motion, reducing air pressure and drawing moist air from neighbouring forests into the openings. The moist air rises, and so generates

*This article and the paper under discussion⁶ were published online on 5 September 2012.



Cover illustration
by Nik Spencer

Editor, *Nature*
Philip Campbell

Publishing
Nick Campbell
Claudia Deasy

Production Editor
Jenny Rooke

Art Editor
Nik Spencer

Sponsorship
Yuki Fujiwara
Yvette Smith
Gerard Preston

Production
Emilia Orviss

Marketing
Elena Woodstock
Hannah Phipps

Editorial Assistant
Anastasia Panoutsou

Now is an exciting time to study the microbes that are associated with the body — the human microbiota. Although the beneficial effects of some of these microbes have been documented for more than a century, only with recent technological advances can the composition of the diverse microbial communities be described and compared within and between individuals. Two major funding initiatives — the US Human Microbiome Project and the European Metagenomics of the Human Intestinal Tract — are well under way, and with the contributions of individual pioneer research groups, we are beginning to understand what constitutes a healthy microbiota and how changes in its structure correlate with and affect health and disease. The more we learn about these microbes, the more we realize that they are influencing many aspects of our physiology.

This Insight focuses on the gut — the body site that is most densely populated with microbes. The greatest challenge in microbiota research is the number and diversity of species in the intestine. Rob Knight and colleagues provide a view on intestinal microbiota using principles derived from macroecology, and discuss how these principles can help to unravel the factors that shape microbial communities and may offer ways to target the microbiota for clinical intervention. Casey Weaver and colleagues review the interplay between the microbiota and the immune system by discussing the mechanisms that maintain a healthy balance, as well as the pathological consequences following perturbation. One of the main functions of gut microbes is to aid nutrient acquisition from otherwise indigestible fibres. Valentina Tremaroli and Fredrik Bäckhed explore the interactions between microbes and metabolic processes, and describe how these exchanges can lead to either a healthy equilibrium or metabolic disease. Much of the research would not have been possible without advances in sequencing technology and analysis of the microbial community genome — the microbiome. George Weinstock reviews the genomic tools that researchers can use to analyse the microbiota.

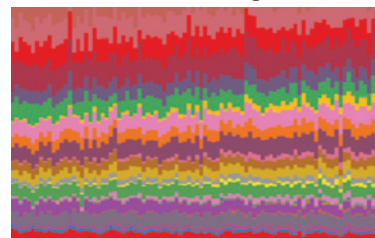
We hope that these Reviews capture some of the most exciting insights into microbiota research. We are pleased to acknowledge the financial support of Yakult Honsha in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content.

Claudia Lupp, Magdalena Skipper & Ursula Weiss
Senior Editors

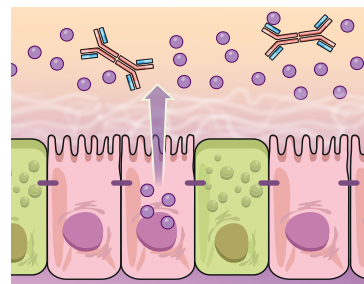
CONTENTS

REVIEWS

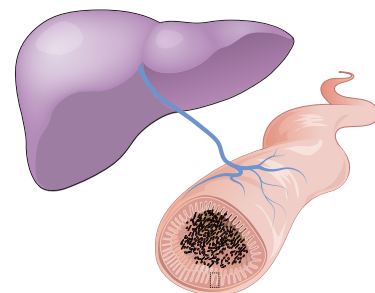
- 220 Diversity, stability and resilience of the human gut microbiota**
Catherine A. Lozupone, Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson & Rob Knight



- 231 Reciprocal interactions of the intestinal microbiota and immune system**
Craig L. Maynard, Charles O. Elson, Robin D. Hatton & Casey T. Weaver



- 242 Functional interactions between the gut microbiota and host metabolism**
Valentina Tremaroli & Fredrik Bäckhed



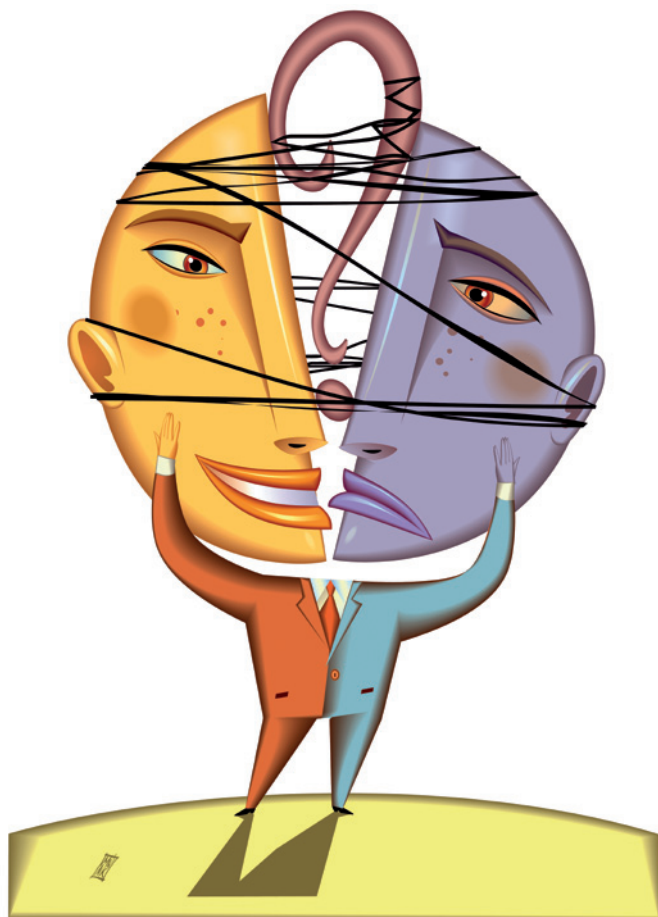
- 250 Genomic approaches to studying the human microbiota**
George M. Weinstock

CAREERS

TURNING POINT Astrobiologist's passion for Mars science creates opportunities **p.328**

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com



faculty members on reviewing applications — 10–20 hours, depending on the size of the pile — and the costs are even higher. No one wants a candidate to turn down an offer.

The joint hiring of academic couples has quadrupled since the 1970s. By 2008, 36% of academics in the United States had a partner who also worked in academia. And the rate was highest among scientists: 54% of male and 83% of female scientists had a partner in academic science, most in a similar field (L. Schiebinger *et al. Dual-Career Academic Couples: What Universities Need to Know*; Stanford University, 2008). With so many female scientists partnered with an academic, strategies to increase the number of women employed as faculty members in science, technology, engineering and mathematics should include a means of addressing dual-career opportunities.

The UNL tackled this question using an Institutional Transformation award from the US National Science Foundation's ADVANCE programme, which aims to improve the representation of women in academic scientific careers. The university came up with a process that should transfer well to other institutions.

FROM THE START

The UNL Dual Career Assistance initiative addresses the two-body problem early in the negotiation and recruiting process. Primary candidates are informed about the programme once they have been shortlisted, but before they are interviewed. They are told that if they have a partner who is qualified for a faculty position, the partner should send a letter of interest, a CV and a statement of research and teaching interests to the ADVANCE office. The director of the ADVANCE programme notifies the appropriate department chair and the college dean of the potential dual-career hire.

This initiative gives the candidate a safe place to disclose their relationship status without making it a factor in deliberations, and provides a relay office to alert all parties. The extra time allows the university to bring the partner to campus for an interview as soon as the offer for the primary candidate is put together. The original search committee is left out of the flow of information, so that it doesn't inappropriately weigh relationship status as a criterion for selection. After all, it may be hard to resist the urge to speculate on how having a partner will affect a candidate's likelihood of accepting an offer. But relationship status, of course, has little to do with the candidate's suitability for the position.

If a candidate with an academic partner is ►

COLUMN

Working together

The right administrative system can ease the job search for dual-career couples, argues **Mary Anne Holmes**.

It is a process common on university campuses around the world. A faculty search committee ploughs through piles of applications, sorts them, evaluates them, brings the best candidates to campus and interviews the top few. It eventually arrives at the perfect person for the job. The head of the department makes the offer — only to discover that the perfect candidate has a partner, and the dreaded two-body problem must be addressed.

It can be a nightmare: the whole search

collapsing as the clock runs out and the committee tries to find a position for the partner. But at the University of Nebraska–Lincoln (UNL), we have developed a means of dealing with the problem to help both candidate and recruiter.

By the time a job offer is made, a substantial amount of time and money has been invested. At the UNL, departments spend between US\$9,000 and \$18,000 simply advertising a position. Factor in the time spent by

► ultimately selected for a faculty position and the partner's target college and department are interested, the partner is flown to Lincoln for an interview. If the interview goes well and faculty members in the target department vote to make an offer to the partner, two offers go forward. If the department is not interested, the process stops and the primary candidate and their partner have to decide whether to come to the university with just one job (we have lost a few candidates this way).

The ADVANCE grant supplies one-quarter of the salary for the partner for up to three years, with the department, college and provost's office each supplying a further quarter. The university's office of research provides start-up funds. This three-year bridge funding buys time for all the administrators to find a permanent funding stream for the partner, which might be provided by a new funding line or the retirement of an existing faculty member.

Sometimes, the primary candidate has a partner who does not want to stay in academia. We can assist with external job searches through our networks and the local chamber of commerce, or the partner can look for a job independently.

The UNL's original ADVANCE grant was for five years, and we anticipated that as many as eight dual-career hires would be made over that time. In fact, after four years we have hired 12 partners.

What happens when the ADVANCE funding runs out? Bridge funding could help. But there are ways, in principle, for universities to institutionalize this process without any grants. One model could be to reserve a proportion of new positions for dual-career hires. The provost or dean could refrain from filling a few positions each year, or could raise an endowment.

Overall, the UNL has had more success with these formal procedures than with informal last-minute requests to accommodate a partner. The programme works in large part because department chairs agree to cooperate: if we accept a partner now, the reasoning goes, another department will help us in the future. Some administrators have expressed concern that a formal dual-career programme will raise the expectations of partners, only for these to be dashed when no position is forthcoming. However, we find that prospective faculty members appreciate the transparency of the process.

Dual-career partnership is not a passing fad. Addressing this need must be an integral part of future universities if we are to attract and keep the best and the brightest. ■

Mary Anne Holmes is a professor of Earth and atmospheric sciences at the University of Nebraska-Lincoln and director of ADVANCE-Nebraska.

TURNING POINT

Alberto Gonzalez Fairén

As a teenager in Spain, Alberto Gonzalez Fairén watched US astronomer Carl Sagan's television series Cosmos, read the book based on the series and knew that he wanted to be a scientist. He is now an astrobiologist at the SETI Institute's Carl Sagan Center for the Study of Life in the Universe in Mountain View, California, and the NASA Ames Research Center in Moffett Field, California. In October, Fairén will collect the Harold C. Urey Prize for outstanding achievement in planetary science from the American Astronomical Society, based in Washington DC.

What was it about *Cosmos* that sparked your interest in science?

Listening to Sagan telling stories about how science is done and how scientists work, and describing the thrill of scientific discovery, I said to myself, "That's what I want to do."

You did a master's degree in genetics. How did you move from that to Martian environments?

It was the other way around. I've been very interested in planetary science and astrobiology since I was a teenager, but there were no university programmes in the field in Spain at the time. Genetics was a distant second interest to me, but I needed to graduate to start my PhD and begin working as a planetary scientist. Astrobiology was starting to become respectable, and I thought that a background in genetics would be appropriate for an astrobiologist interested in Martian environments.

What are your plans for the future?

I'm leaving the SETI Institute because 'soft money' — working solely off grants — is no longer an option for me. When you have a family, you need some kind of stability. I've been offered two academic positions in the United States, and I've won a starting grant from the European Research Council. If I accept one of the offers here, I will stay in US Mars science, which is very attractive right now, especially after the success with the Curiosity rover landing. If I go back to Europe, I will have substantial funding to launch my own group, which seems the logical next step.

Describe your biggest career milestone.

Carbonates are minerals produced in water; the fact that none had been found on Mars had long been used to argue that there were never oceans on Mars. But my adviser told me about a river in Spain that was very acidic; I knew that carbonates do not form in acidic water. I was two years into my PhD and proposed that



oceans on Mars were also acidic and that this acidity inhibited the formation of carbonates. A team including me and my adviser submitted a paper on the subject to *Nature* on 7 January 2004. NASA's rover Opportunity landed on Mars days later — and found direct evidence for such acidic environments. We had accurately anticipated the rover's results. Our paper was published on 23 September 2004, and was central to the completion of my PhD.

Does the Urey prize have a special meaning for you?

It won't help me to get grants or positions, but it is a high personal honour. There is a long list of names of very, very, very important people in planetary science who have won this award.

Has it been difficult to adjust to living and working in the United States?

Not really. Life in California is not so different from life in Spain, and I already had a good command of English. The biggest difference is that the United States rewards effort and merit.

Do you have any regrets about your career?

No. Some time ago I realized that my errors were necessary to bring me to where I am now. I didn't begin to pursue my PhD until I was 30, and while I was away from science I worked as a bartender, a taxi driver, a film-rating classifier and a teacher, among other things. But I got my PhD in 2006 and started working for NASA, and that was a turning point in my career. It offered me the freedom and resources to pursue my own investigations. My career would have been totally different without the NASA experience. ■

INTERVIEW BY KAREN KAPLAN

COMMUNICANT

The ties that bind.

BY JOHN GILBEY

We'd been dating for years, but we were still tiptoeing around the question that faces many couples in our situation. It was, we agreed very late one night, a matter of commitment and trust — and not something to take on lightly.

In retrospect, it was probably me that blocked the discussion from going any further which — I realized far too late — may well have been why Helen was so enthusiastic about the Silicon Valley job. It was a wonderful opportunity: three months with a bio-data start-up company in Mountain View getting them up to speed on the new techniques that she had just published.

Sure, we vowed to stay in touch, but as her work ramped up and the schedule began to bite, our conversations all but petered out. Glumly, I began to understand just what a mistake I'd made. I had, I reasoned, completely blown it — so I was really surprised when she wanted to see me the first evening after she'd got back from her trip. She had, she said, something to tell me.

We agreed to meet for dinner at the Tandoori place near the railway station — where we'd had our first date. I got there early and sat nervously scrunching poppadoms and drinking too much cold Indian lager while mentally preparing my speech of apology.

She looked fantastically alive. The Californian sun had tinted her hair slightly auburn and her skin had acquired a fresh glow of health and well-being. Helen smiled at me — the smile of someone who has important news that they are dying to share. I put two and two together and asked her if she was pregnant. Her laugh came as a relief to me, although the loud whoop I gave gained me some baleful stares from the other diners. So, if not that — then what?

Helen turned slightly in her chair and scooped her hair aside with one hand. There, at the base of her skull, was the thing we had spent so long discussing: a small, dimpled disk of alloy, anodized a fashionable lime green, laser engraved with the tiny characters

defining her range of MAC addresses. Helen had got a Connection.

I can't remember everything I said, but I know I used "betrayal" a few times — and "selfish", and possibly a few worse epithets. Helen sat serenely until I'd finished my rant, then she smiled away the waiters who were trying to come to her rescue.



She looked at me with something that I hoped was compassion — but it was the kind of look you give to a grumpy child who doesn't understand the adult world.

"I know we said we'd wait — but when I got to the lab and saw how much synergy everyone got from being Connected, I just couldn't resist. They were really kind, but I could see that I wouldn't be able to work there effectively without it. It's more than that though — it gives a wonderful sense of... well, belonging and understanding."

I asked — testily — for an example, and Helen thought for a moment. "Well, things like knowing that the X32 bus is running nine minutes late, that the guy over there in the corner is really excited about his new media job — he is here to celebrate with his friends but they are still on the bus. He thinks I'm cute, by the way — he especially likes my... Uh, never mind. Oh, and it's going to rain later — so I'd like to be home by ten... There is just so much richness and texture; I thought you, of all people, would understand." Just because I'm a

software developer doesn't mean I have to like the stuff, I thought gloomily. Having worked in the industry for years, I know just what happens when you are working against a wildly optimistic deadline with a microscopic budget — and that isn't the sort of code I want plumbed permanently into my nervous system, even if it does mean I

can read my e-mail while walking down the street. I considered remonstrating with her about the lack of adequate testing protocols, the myriad unknown risks of interactions and privacy — don't get me started on privacy — but they were all old, stale arguments.

Helen now had a golden world — a world without me — with the millions of other Connected folk who obviously shared startlingly intimate details of her life. Nothing I could do or say would change her delight in this new set of senses. I have to admit, though, that I was curious about what it felt like when the software was running, so I asked her.

She wrinkled her nose, then her eyes defocused slightly while she reviewed her symbiotic systems and tried to put words to the sensations. "Kind of nice, like wearing a cosy pair of slippers — or walking with a friend. But the best bit is when an application gets upgraded, there's a sort of tingly rush of anticipation like you get on Christmas Eve when you're a kid."

Her cheeks flushed slightly in a way that I recognized from a much more private moment. "In fact," she said quietly, "the core system is updating itself right now."

I waited expectantly for a moment, and then slowly became aware of the lack of movement in the space around me. Across the table, Helen was rigid — with her pupils slowly dilating. As I watched with mounting horror, a small bead of saliva slid from one side of her mouth. Then the screaming started: sharp, guttural and inhuman — but it was a long time before I realized it was coming from me. ■

John Gilbey is currently wrestling with his first science-fiction novel, Contextual, and fears that it may be gaining the upper hand.

JACEY

Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling

Marcus Dühren-von Minden^{1*}, Rudolf Übelhart^{1,2*}, Dunja Schneider^{1*}, Thomas Wossning¹, Martina P. Bach¹, Maike Buchner³, Daniel Hofmann¹, Elena Surova^{1,2}, Marie Follo³, Fabian Köhler¹, Hedda Wardemann⁴, Katja Zirlik³, Hendrik Veelken⁵ & Hassan Jumaa^{1,6}

B-cell antigen receptor (BCR) expression is an important feature of chronic lymphocytic leukaemia (CLL), one of the most prevalent B-cell neoplasias in Western countries¹. The presence of stereotyped and quasi-identical BCRs in different CLL patients suggests that recognition of specific antigens might drive CLL pathogenesis. Here we show that, in contrast to other B-cell neoplasias, CLL-derived BCRs induce antigen-independent cell-autonomous signalling, which is dependent on the heavy-chain complementarity-determining region (HCDR3) and an internal epitope of the BCR. Indeed, transferring the HCDR3 of a CLL-derived BCR provides autonomous signalling capacity to a non-autonomously active BCR, whereas mutations in the internal epitope abolish this capacity. Because BCR expression was required for the binding of secreted CLL-derived BCRs to target cells, and mutations in the internal epitope reduced this binding, our results indicate a new model for CLL pathogenesis, with cell-autonomous antigen-independent signalling as a crucial pathogenic mechanism.

CLL cases can be divided into two groups based on the mutational status of the immunoglobulin genes. Notably, in B cells from ~1% of both groups, unmutated CLL and mutated CLL cases express almost identical BCRs, and in 30% of cases the BCRs can be classified into certain stereotypes based on structural similarities and the HCDR3 (refs 2–5). This restricted BCR repertoire has been interpreted to be the result of a BCR-driven selection mechanism initiated by specific antigens that may promote the expansion of the respective CLL clone^{6,7}.

We have recently reported functional similarity between the pre-BCR and polyreactive BCRs, and have demonstrated that both activate cell-autonomous signalling in pre-B cells⁸. Furthermore, we have shown that the pre-BCR induces ligand-independent cell-autonomous signalling by surrogate light-chain-mediated binding to an intrinsic pre-BCR glycosylation site⁹. CLL-derived BCRs are also often polyreactive^{10,11}, suggesting that, in analogy to the pre-BCR, CLL BCRs might activate ligand-independent autonomous signalling through the binding of BCR-intrinsic motifs. We tested this hypothesis by comparing the signalling capacity of randomly chosen BCRs from different B-cell lymphomas in triple knockout (TKO) cells that do not express endogenous pre-BCRs or BCRs owing to inactivation of RAG2 and $\lambda 5$ genes. An additional deficiency in SLP65 (also known as BLNK) expression makes TKO cells an ideal reconstitution system to test receptor function by transfection of receptor components and a tamoxifen-inducible ERT2–SLP65 fusion protein¹². As shown previously⁸, the pre-BCR and a polyreactive human BCR derived from an early bone marrow B cell (BCR62) activate Ca^{2+} mobilization autonomously without additional ligands, whereas a non-polyreactive human BCR derived from a selected mature B cell (BCR53) requires additional crosslinking (Supplementary Fig. 1a and data not shown). Similar to BCR53, we found that BCRs derived from multiple myeloma

($n = 4$), marginal zone lymphoma ($n = 3$), mantle cell lymphoma ($n = 4$) and follicular lymphoma ($n = 4$) required crosslinking with anti-BCR antibody in addition to SLP65 activation to induce Ca^{2+} flux (Fig. 1, Supplementary Fig. 1b, c and Supplementary Table 1; data not shown). By contrast, all tested ($n = 17$) CLL-derived BCRs, including those containing typical CLL variable heavy chain (VH) elements (for example, IGHV4-34, IGHV3-7, IGHV3-23, IGHV3-30 and IGHV1-69)³ and stereotyped HCDR3s, induced autonomous signalling on SLP65 activation (Fig. 1, Supplementary Fig. 2a and Supplementary

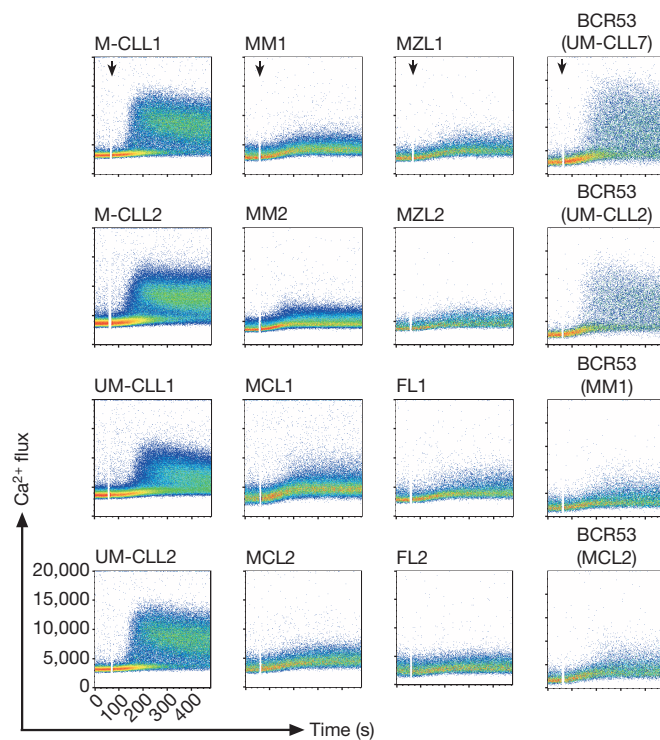


Figure 1 | CLL-derived BCRs possess autonomous signalling capacity. Fluorescence-activated cell sorting (FACS) analyses of Ca^{2+} flux after the addition of 4-hydroxytamoxifen (4-OHT) to TKO cells expressing the ERT2–SLP65 fusion protein and four representative CLL-derived BCRs (left column), two representative BCRs of multiple myeloma (MM), mantle cell lymphoma (MCL), marginal zone lymphoma (MZL) and follicular lymphoma (FL) (middle two columns) and BCR53 variants containing different HCDR3s of representative CLL, multiple myeloma or mantle cell lymphoma receptors (right column). Addition of the stimulus is indicated by black arrows. M, mutated; UM, unmutated.

¹Centre for Biological Signaling Studies (BIOSS), Albert-Ludwigs Universität Freiburg, 79104 Freiburg, Germany. ²Spemann Graduate School of Biology and Medicine, Albert-Ludwigs Universität Freiburg, 79104 Freiburg, Germany. ³Department of Hematology/Oncology, University Medical Center Freiburg, 79106 Freiburg, Germany. ⁴Max Planck Research Group Molecular Immunology, Max Planck Institute for Infection Biology, 10117 Berlin, Germany. ⁵Department of Hematology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands. ⁶Department of Molecular Immunology, Faculty of Biology, Albert Ludwigs University of Freiburg and Max Planck Institute of Immunobiology and Epigenetics, 79108 Freiburg, Germany.

*These authors contributed equally to this work.

Table 1; data not shown). No obvious difference was found between mutated ($n = 10$) and unmutated ($n = 7$) CLL BCRs (Fig. 1 and data not shown). Importantly, BCRs derived from healthy donors who possess CLL-typical VH genes such as *IGHV1-69* or *IGHV4-34* did not display autonomous signalling (Supplementary Fig. 2b).

To investigate whether autonomous signalling is also involved in lymphoproliferation in animal models of CLL, we examined TCL1 mice, which express a transgene for human T-cell leukaemia 1 (*TCL1*) under the control of the VH promoter and the Eμ enhancer¹³. *TCL1* is highly expressed in unmutated CLL¹⁴ and *TCL1* transgenic mice show clonal expansion of CD5⁺ IgM⁺ B cells bearing unmutated BCRs^{13,15}. To test whether transformed clones derived from *TCL1* mice express autonomously active BCRs, we studied four different BCRs from *TCL1* mice in the TKO reconstitution system. All four *TCL1*-derived BCRs showed autonomous Ca²⁺ flux similar to the human CLL BCRs and the murine pre-BCR, whereas a 4-hydroxy-3-iodo-5-nitrophenylacetyl (NIP)-specific murine BCR (B1-8) showed no autonomous signalling (Supplementary Fig. 3a, b). Similar to human B-cell CLL, *TCL1*-transformed murine B cells are CD5⁺. Therefore, we tested whether autonomous signalling is a characteristic feature of murine B1a B cells that are typically CD5⁺ and IgM⁺. Interestingly, BCRs from B1a B cells did not induce autonomous signalling, indicating that this capacity is a unique feature of leukaemia-derived BCRs (Supplementary Fig. 3c). Together, our data show that, in contrast to all other B-cell lymphomas tested, BCRs derived from CLL patients and *TCL1* mice possess the unique capacity to induce autonomous signalling.

The bias in VH gene usage and the high proportion of stereotyped HCDR3s in CLL suggest that HCDR3 has a critical role in the development of this disease^{2,3,5}. To test whether transferring HCDR3 from a CLL-derived BCR transfers autonomous signalling capacity to a non-autonomously active BCR, we replaced the HCDR3 of BCR53 with three different CLL-derived HCDR3s. As a negative control, we replaced the HCDR3 of BCR53 with the HCDR3s of a multiple myeloma BCR and a mantle cell lymphoma BCR that contained the same VH gene as the chosen CLL BCRs. Although all modified BCR53 derivatives were expressed on the cell surface, only those containing a CLL-derived HCDR3 acquired the capacity for autonomous signalling (Fig. 1 and Supplementary Fig. 1; data not shown).

Our previous results indicated that the ability to induce autonomous Ca²⁺ flux was associated with BCR polyreactivity⁸. Because most secreted CLL-derived BCRs were reported to be polyreactive¹⁰, we tested the chosen CLL BCRs for polyreactivity. Only three out of six randomly chosen CLL-derived BCRs—two unmutated and one mutated—showed polyreactive potential by binding to more than two antigens in enzyme-linked immunosorbent assays (ELISA) and showing cytoplasmic staining of HEp-2 slides in indirect immunofluorescence assays (Supplementary Fig. 4a, b). Furthermore, only one out of four tested *TCL1* BCRs (*TCL1* no. 3) showed polyreactivity in the same assays (Supplementary Fig. 4c, d). The frequent lack of polyreactivity suggests that other mechanisms induce the autonomous signalling of CLL-derived BCRs. To show that specific antigens on adjacent cells are not required for the activation of CLL BCRs, we examined autonomous Ca²⁺ signalling in isolated, single cells. We detected autonomous Ca²⁺ flux in single cells expressing a CLL-derived BCR, whereas additional stimulation with anti-BCR antibodies was required for activation of single cells expressing the BCR53 control (Fig. 2a, b).

The lack of polyreactivity suggests that antigen-independent signalling might be achieved by HCDR3-dependent binding of BCR-intrinsic motifs, suggesting mutual BCR recognition on CLL cells. Indeed, secreted CLL-derived BCRs bound to the surface of TKO cells expressing a CLL BCR or the control BCR53, although they failed to bind to TKO cells that did not express a receptor (Fig. 3a, b and Supplementary Fig. 5a). This suggests that the CLL-derived BCRs might interact with intrinsic BCR structures and that this interaction

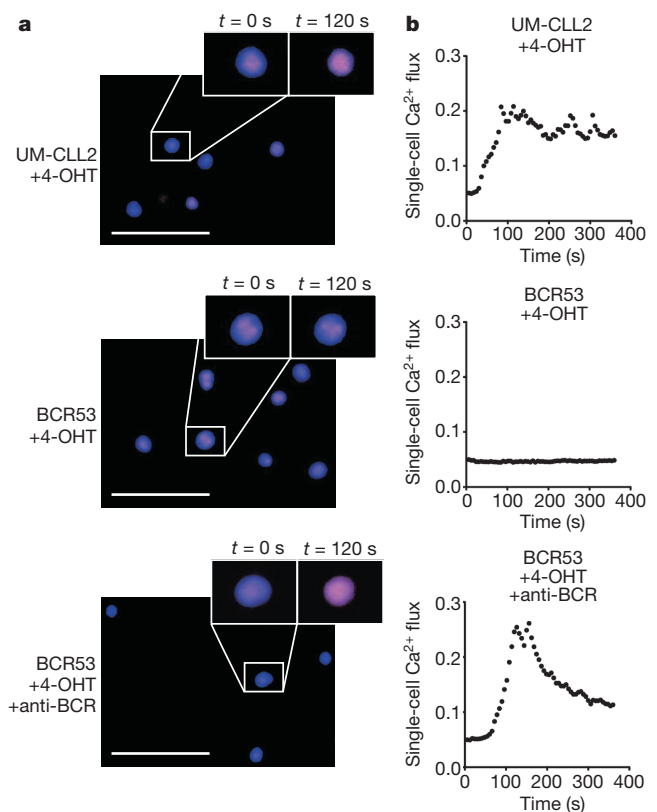


Figure 2 | CLL BCRs induce cell-autonomous signalling. **a**, Live-cell imaging of single TKO cells expressing ERT2-SLP65 and unmutated CLL2 or BCR53 after induction with 4-OHT. Images show merged channels of Indo-1 unbound (blue) and Indo-1 bound (violet). Magnification is $\times 40$; scale bars, 25 μm . **b**, Ca²⁺ flux of representative single cells from **a**. Kinetics are representative of at least ten single cells.

could mediate antigen-independent stimulation if it takes place within the membrane of the same cell. Using a phage-display approach, a recent study identified different peptide motifs that bind to primary CLL cells in a HCDR3-dependent manner, and one of these motifs showed specific binding to each of the tested CLL samples¹⁶. Through sequence analysis we found that this motif possesses homology to a conserved epitope in the framework region 2 (FR2) of VH domains (Supplementary Fig. 5c). To test whether this region contains an internal epitope important for autonomous signalling, we generated point mutations of the amino acids Val 37 and Arg 38 within this motif. As a control we mutated a conserved arginine (Arg 43) outside of the motif. Our results show that mutating Val 37 or Arg 38, but not Arg 43, abolishes autonomous signalling of CLL BCRs, whereas stimulation with anti-BCR antibodies show normal Ca²⁺ flux for all mutants (Fig. 3c, d and Supplementary Fig. 5b, d; data not shown). As the introduced mutations are located outside of the antigen-recognition site of the BCR, these data indicate that the observed autonomous signalling is not induced by BCR interaction with external antigens in the environment, because such antigens should still be able to bind to and stimulate the mutated CLL BCR. To confirm this view, we mutated Val 37 and Lys 38 in the NIP-specific B1-8 BCR and found that the resulting mutant BCRs show normal antigen-dependent stimulation after incubation with NIP-bovine serum albumin (Fig. 3c and Supplementary Fig. 5b). Thus, these data indicate that the loss of autonomous signalling was caused by the inability of the CLL BCR to bind to the mutated intrinsic motif. In agreement with this, the binding of secreted CLL BCRs to TKO cells was reduced when the cells expressed a BCR with a mutated intrinsic motif (Fig. 3e). Together, our results suggest that an intrinsic BCR motif and HCDR3 mediate autonomous BCR stimulation in CLL-derived BCRs.

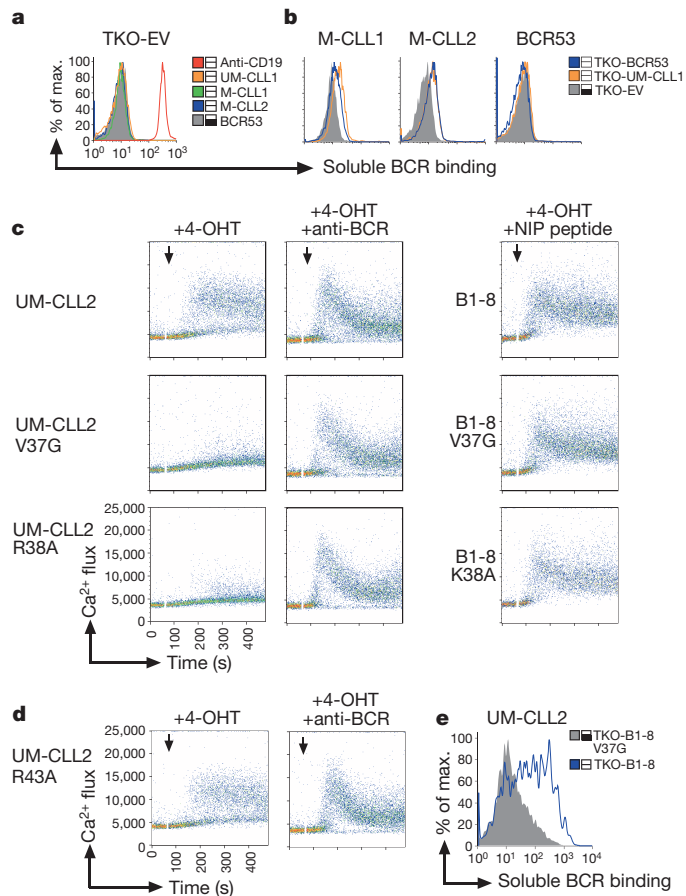


Figure 3 | Autonomous signalling of CLL-derived BCRs is mediated by the recognition of a BCR-intrinsic epitope. **a**, Surface-binding profiles of different secreted CLL BCRs on TKO cells expressing empty vector (EV). Secreted BCR53 and anti-CD19 antibody were used as negative and positive controls, respectively. The amount of secreted BCR used was >100-fold higher than for the anti-CD19 antibody. **b**, Surface-binding profiles of secreted M-CLL1, M-CLL2 and BCR53 on TKO cells expressing the indicated BCRs. **c**, **d**, FACS analysis for autonomous Ca^{2+} flux after addition of the indicated stimuli in TKO cells expressing the ERT2-SLP65 fusion protein. **c**, The UM-CLL2 mutants V37G and R38A (left) or the corresponding B1-8 mutants (right). **d**, The UM-CLL2 R43A mutant. Arrows indicate addition of the stimuli. **e**, Surface-binding profile of UM-CLL2 on TKO cells expressing B1-8 or B1-8 V37G.

To show that autonomous signalling can also be detected in primary CLL samples, we monitored Ca^{2+} flux in primary B cells from CLL patients and compared them to primary peripheral B cells from healthy donors or patients with other B-cell neoplasias. Because SLP65 function is constitutively active in primary cells, a synchronous start of autonomous signalling as in the TKO cells cannot be achieved in primary cells. However, constitutive autonomous signalling suggests that a proportion of the primary cells show increased Ca^{2+} flux. In fact, our experiments show that primary CLL samples show elevated Ca^{2+} mobilization compared with the control cells. Moreover, this signalling capacity required BCR function, as inhibition of the protein tyrosine kinase SYK, which transmits BCR signals¹⁷, prevented this activity (Fig. 4 and Supplementary Fig. 6a). Similarly, monitoring Ca^{2+} mobilization in single cells showed periodic SYK-dependent signals in isolated primary CLL B cells but not in control cells (Fig. 4 and Supplementary Fig. 6b). As survival of CLL B cells is associated with nuclear-factor κB (NF- κB) activation¹⁸, we investigated the expression of NF- κB -regulated survival genes¹⁸ in primary CLL cells and found that inhibition of autonomous BCR signalling by inhibition of SYK led to a reduction in the expression of MYC and XIAP (Supplementary Fig. 6c).

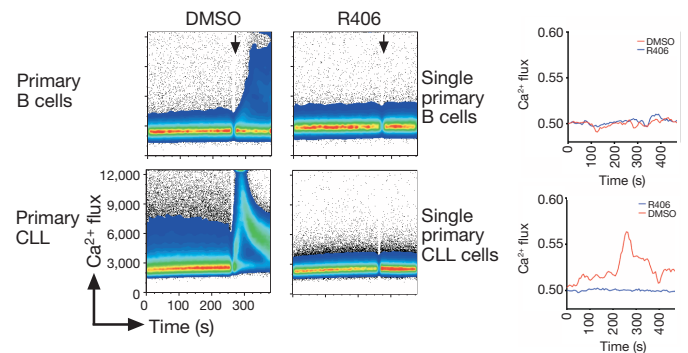


Figure 4 | Cell-autonomous Ca^{2+} signalling is elevated in primary CLL B cells. Ca^{2+} flux in primary peripheral B cells from a healthy donor or a CLL patient with or without R406 treatment (left panels; addition of an anti-BCR stimulus as a control for cell viability and Indo loading is indicated by a black arrow) and Ca^{2+} flux in single cells from a healthy donor or a CLL patient (right panels). DMSO, dimethylsulphoxide.

BCR expression is commonly used as a diagnostic marker for the characterization of CLL patients. The restricted BCR repertoire of both mutated and unmutated CLL suggests a potentially important role for BCR signalling in the transformation and expansion of CLL B cells. The restricted BCR repertoire and the observed polyautoactivity suggested that CLL BCRs might recognize extrinsic autoantigens, thereby inducing the selective expansion of the respective B cells. Therefore, blocking such autoantigens or blocking BCR function by inhibiting downstream signalling factors, such as the tyrosine kinase SYK, are considered as potential therapies for CLL^{19–22}.

However, our data indicate that CLL BCRs induce cell-autonomous signalling independent of extrinsic antigens and identify an intrinsic motif that might provide new targets for the therapy of CLL by interfering with autonomous BCR stimulation. It is conceivable that additional intrinsic epitopes exist that activate the autonomous signalling of distinct CLL BCRs. In this scenario, the HCDR3-driven interaction with distinct epitopes on nearby BCRs is sufficient for the induction of autonomous signalling on CLL cells. In agreement with this, HCDR3 sequences derived from CLL BCRs were able to transfer the autonomous signalling capacity to other receptors in our experiments. Presumably, only certain HCDR3 sequences have this unique capacity, which might explain the restricted repertoire and the requirement for certain HCDR3 regions in CLL. Thus, the presence of distinct intrinsic epitopes may explain the observed structural similarities of CLL BCRs, which are probably selected on the basis of the ability of HCDR3 to interact with such intrinsic structures. It should be noted, however, that the presence of intrinsic epitopes and their role in autonomous signalling does not rule out that for certain cases, extrinsic antigens are involved in the pathogenesis of CLL. This may further potentiate the basal activity of CLL BCRs. Our study presents a new mechanism for the pathogenesis of CLL and suggests that investigating intrinsic BCR-associated epitopes may lead to the development of new therapeutic approaches based on the inhibition of the autonomous BCR activation in CLL cells.

METHODS SUMMARY

Details of lymphoma material, clinical data, mice strain data, cells and cell-culture conditions, polyreactivity assays, plasmids and retroviral transduction, flow cytometry and Ca^{2+} flux analysis, single-cell Ca^{2+} flux analysis, NF- κB -activity assays, reverse-transcriptase PCR and statistical analysis are provided in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 28 November 2011; accepted 11 June 2012.

Published online 12 August 2012.

1. Morton, L. M. *et al.* Lymphoma incidence patterns by WHO subtype in the United States, 1992–2001. *Blood* **107**, 265–276 (2006).

2. Ghiotto, F. *et al.* Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. *J. Clin. Invest.* **113**, 1008–1016 (2004).
3. Stamatopoulos, K. *et al.* Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. *Blood* **109**, 259–270 (2007).
4. Bikos, V. *et al.* Over 30% of patients with splenic marginal zone lymphoma express the same immunoglobulin heavy variable gene: ontogenetic implications. *Leukemia* **26**, 1638–1646 (2012).
5. Widhopf, G. F. II *et al.* Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. *Blood* **104**, 2499–2504 (2004).
6. Ghia, P., Chiorazzi, N. & Stamatopoulos, K. Microenvironmental influences in chronic lymphocytic leukaemia: the role of antigen stimulation. *J. Intern. Med.* **264**, 549–562 (2008).
7. Stevenson, F. K., Krysov, S., Davies, A. J., Steele, A. J. & Packham, G. B-cell receptor signaling in chronic lymphocytic leukemia. *Blood* **118**, 4313–4320 (2011).
8. Köhler, F. *et al.* Autoreactive B cell receptors mimic autonomous pre-B cell receptor signaling and induce proliferation of early B cells. *Immunity* **29**, 912–921 (2008).
9. Übelhart, R. *et al.* N-linked glycosylation selectively regulates autonomous precursor BCR function. *Nature Immunol.* **11**, 759–765 (2010).
10. Hervé, M. *et al.* Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. *J. Clin. Invest.* **115**, 1636–1643 (2005).
11. Sthoeger, Z. M. *et al.* Production of autoantibodies by CD5-expressing B lymphocytes from patients with chronic lymphocytic leukemia. *J. Exp. Med.* **169**, 255–268 (1989).
12. Meixlsperger, S. *et al.* Conventional light chains inhibit the autonomous signaling capacity of the B cell receptor. *Immunity* **26**, 323–333 (2007).
13. Bichi, R. *et al.* Human chronic lymphocytic leukemia modeled in mouse by targeted *TCL1* expression. *Proc. Natl Acad. Sci. USA* **99**, 6955–6960 (2002).
14. Herling, M. *et al.* High *TCL1* levels are a marker of B-cell receptor pathway responsiveness and adverse outcome in chronic lymphocytic leukemia. *Blood* **114**, 4675–4686 (2009).
15. Yan, X. J. *et al.* B cell receptors in *TCL1* transgenic mice resemble those of aggressive, treatment-resistant human chronic lymphocytic leukemia. *Proc. Natl Acad. Sci. USA* **103**, 11713–11718 (2006).
16. Binder, M. *et al.* B-cell receptor epitope recognition correlates with the clinical course of chronic lymphocytic leukemia. *Cancer* **117**, 1891–1900 (2011).
17. Kulathu, Y., Grothe, G. & Reth, M. Autoinhibition and adapter function of Syk. *Immunol. Rev.* **232**, 286–299 (2009).
18. Herishanu, Y. *et al.* The lymph node microenvironment promotes B-cell receptor signaling, NF- κ B activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood* **117**, 563–574 (2011).
19. Seiler, T. *et al.* Characterization of structurally defined epitopes recognized by monoclonal antibodies produced by chronic lymphocytic leukemia B cells. *Blood* **114**, 3615–3624 (2009).
20. Buchner, M. *et al.* Spleen tyrosine kinase inhibition prevents chemokine- and integrin-mediated stromal protective effects in chronic lymphocytic leukemia. *Blood* **115**, 4497–4506 (2010).
21. Suljagic, M. *et al.* The Syk inhibitor fostamatinib disodium (R788) inhibits tumor growth in the E μ -*TCL1* transgenic mouse model of CLL by blocking antigen-dependent B-cell receptor signaling. *Blood* **116**, 4894–4905 (2010).
22. Friedberg, J. W. *et al.* Inhibition of Syk with fostamatinib disodium has significant clinical activity in non-Hodgkin lymphoma and chronic lymphocytic leukemia. *Blood* **115**, 2578–2585 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Croce for providing the *TCL1* mice, M. Reth and P. Nielsen for scientific discussion and reading the manuscript, C. Wehr for help cloning *TCL1*-derived BCRs, A. Ott, A. Würch, S. Hopitz and B. Wehrle for help with FACS experiments and cloning, and D. Pfeifer and M. Pantic for providing clinical data. This work was supported by the Deutsche Krebshilfe (Project 108935), the Deutsche Forschungsgemeinschaft (SFB746 and JU 463/2-1) and the Excellence Initiative of the German Federal and State Governments (GSC-4, Spemann Graduate School).

Author Contributions M.D.-v.M., R.Ü., D.S., F.K. and M.P.B. conducted cloning of BCRs. M.D.-v.M., R.Ü. and D.S. performed measurements of BCR surface expression and calcium flux and performed soluble BCR-binding assays. M.D.-v.M. and T.W. performed quantitative reverse-transcriptase PCR experiments. M.B. performed the NF- κ B activity ELISA. E.S. analysed V μ region sequences from patient samples. M.D.-v.M. and M.F. performed single-cell Ca²⁺ measurements. D.H. and H.W. tested the polyreactivity of CLL and *TCL1* receptors. K.Z. and H.V. conducted the characterization of lymphoma patients and provided samples. M.D.-v.M., T.W., F.K. and H.J. wrote the manuscript. H.J. designed experiments. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.J. (jumaa@immunbio.mpg.de).

METHODS

Lymphoma material and clinical data. CLL B-cell samples were obtained from the peripheral blood of patients at the University Medical Center Freiburg. Data for IGVH status, ZAP70 expression and time to first treatment were taken from the clinical records. The local ethics committee approved patient sampling and all patients gave informed consent.

Mice. Eμ-TCL1 mice (provided by C. Croce) were described previously¹³. Mice were bred in the animal facility of the Max Planck Institute of Immunobiology and Epigenetics under specific pathogen-free conditions. Animal experiments were done in compliance with the guidelines of the German law.

Cells and cell-culture conditions. HEK293T, Phoenix and TKO cells were cultured in Iscove's Medium (Biocrom AG) containing 5% FCS (PAN-Biotech), 10 mM L-glutamine (Gibco) and 100 U ml⁻¹ penicillin/streptomycin (Gibco). For TKO cultures, the medium was supplemented with 50 mM 2-mercaptoethanol (Gibco) and supernatant of J558L mouse plasmacytoma cells stably transfected with a murine IL-7 expression vector.

Primary peripheral B cells from CLL patients and healthy donors were cultured in RPMI containing 10% FCS, 10 mM L-glutamine and 100 U ml⁻¹ penicillin/streptomycin for 24 h. For the inhibition of SYK, cells were then treated with 4 μM R406 (Rigel) for 1 h.

ELISA and indirect immunofluorescence assays. ELISA assays were performed using microtitre plates coated with 10 μg ml⁻¹ of single-stranded DNA, double-stranded DNA (ORGENTEC), lipopolysaccharide or human insulin (Sigma-Aldrich). CLL antibodies were produced as described previously²³ and purified antibodies were used at 1 mg ml⁻¹ and three serial dilutions of 1:4 in PBS. All ELISA assays were developed with horseradish-peroxidase-labelled goat anti-human IgG (BioSource). $D_{405\text{ nm}}$ was measured using a microtitre plate reader (SpectraMax 190).

Indirect immunofluorescence assays were done as previously described²⁴. In brief, diagnostic HEP-2 slides were incubated with 15 μl purified CLL antibodies (1 mg ml⁻¹). After washing, a Cy3-conjugated anti-human IgG1 antibody (Jackson ImmunoResearch Laboratories) was used to detect binding. Nuclear staining was performed with 4',6-diamidino-2-phenylindole (DAPI).

Plasmids and retroviral transduction. ERT2-SLP65 (ref. 12) was expressed using a vector containing tdTomato as expression control (a gift from R. Y. Tsien). Immunoglobulin heavy chain and light chain were expressed using the biomolecular fluorescence complementation vector system described previously⁸.

CLL-derived BCRs were amplified from patient samples by anchor-PCR using poly-G-tailed complementary DNA and a poly-C-containing primer as described^{25,26}. Human VH domains were fused to the murine constant region (μCH) by PCR, whereas the complete human light chains were used. Other lymphoma-derived BCRs were amplified from sequencing vectors containing heavy-chain or light-chain sequences of multiple myeloma, mantle cell lymphoma, marginal zone lymphoma and follicular lymphoma (CellGenix). For virus production, phoenix cells were transfected using FuGeneHD as

recommended by the manufacturer's protocol. Supernatants were collected 48 h after transfection and used for transduction of TKO cells as described²⁷.

Flow cytometry and Ca²⁺ flux analysis. For FACS analyses, TKO cells were stained with goat anti-mouse IgM (Southern Biotech), goat anti-human κ light chain and goat anti-human λ light chain (Southern Biotech).

Ca²⁺ analyses were performed as described²⁸. In brief, 1 × 10⁶ freshly transduced TKO cells expressing ERT2-SLP65 were loaded with Indo-1 (Invitrogen) using Pluronic (Invitrogen). Induction of ERT2-SLP65 was performed by the addition of 2 μM 4-OHT (Sigma-Aldrich). Goat anti-mouse IgM (10 μg ml⁻¹; Southern Biotech) was used for crosslinking of the BCR. For the inhibition of SYK, cells were pre-treated with 4 μM R406 (Rigel) for 1 h.

Single-cell Ca²⁺ flux analysis. TKO cells expressing ERT2-SLP65 and BCR53 or unmutated CLL2 were prepared as described²⁸ and seeded onto coated eight-well μ-slides (ibidi) using BD Cell-Tak (BD Biosciences). Live-cell imaging of single cells was performed using a ScañR screening station on an IX81 inverse microscope (Olympus). Measurements were taken at 37 °C and at time intervals of 6 min after induction with 4-OHT or 4-OHT and goat anti-mouse IgM (Southern Biotech) or goat anti-human IgM. For the inhibition of SYK, cells were pre-treated with 4 μM R406 (Rigel) for 1 h.

NF-κB activity assay. NF-κB activity was measured using the TransAM NF-κB p50 transcription factor assay kit (Active Motif). In brief, peripheral blood B cells from CLL patients were treated with 4 μM R406 for 1 h or were left untreated. Total cellular lysates were applied to 96-well plates coated with oligonucleotides containing the NF-κB consensus sequence (5'-GGGACTTTC-3') and NF-κB activity was calculated by measuring p50 binding using a specific antibody.

Quantitative reverse-transcriptase PCR. Total RNA was extracted, reverse-transcribed using poly-dT primers and amplified using MYC-specific primers and XIAP-specific primers designed by Universal ProbeLibrary Assay Design Center (Roche Applied Science). In addition, the amount of actin transcripts was quantified in all samples as an internal control. SYBR green was used in the detection method.

Statistical analysis. For statistical analysis, GraphPad Prism software was used.

23. Wardemann, H. *et al.* Predominant autoantibody production by early human B cell precursors. *Science* **301**, 1374–1377 (2003).
24. Tiller, T. *et al.* Development of self-reactive germinal center B cells and plasma cells in autoimmune FcγRIIB-deficient mice. *J. Exp. Med.* **207**, 2767–2778 (2010).
25. Bertinetti, C. *et al.* Cloning of idiotype immunoglobulin genes in B cell lymphomas by anchored PCR and production of individual recombinant idiotype vaccines in *Escherichia coli*. *Eur. J. Haematol.* **77**, 395–402 (2006).
26. Osterroth, F. *et al.* Rapid expression cloning of human immunoglobulin Fab fragments for the analysis of antigen specificity of B cell lymphomas and anti-idiotype lymphoma vaccination. *J. Immunol. Methods* **229**, 141–153 (1999).
27. Su, Y. W. & Jumaa, H. LAT links the Pre-BCR to calcium signaling. *Immunity* **19**, 295–305 (2003).
28. Storch, B., Meixlsperger, S. & Jumaa, H. The Ig-α ITAM is required for efficient differentiation but not proliferation of pre-B cells. *Eur. J. Immunol.* **37**, 252–260 (2007).

RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions

David Vilchez¹, Ianessa Morantte¹, Zheng Liu¹, Peter M. Douglas¹, Carsten Merkwirth¹, Ana P. C. Rodrigues^{2†}, Gerard Manning^{2†} & Andrew Dillin¹

Organisms that protect their germ-cell lineages from damage often do so at considerable cost: limited metabolic resources become partitioned away from maintenance of the soma, leaving the ageing somatic tissues to navigate survival amid an environment containing damaged and poorly functioning proteins. Historically, experimental paradigms that limit reproductive investment result in lifespan extension. We proposed that germline-deficient animals might exhibit heightened protection from proteotoxic stressors in somatic tissues. We find that the forced re-investment of resources from the germ line to the soma in *Caenorhabditis elegans* results in elevated somatic proteasome activity, clearance of damaged proteins and increased longevity. This activity is associated with increased expression of *rpn-6*, a subunit of the 19S proteasome, by the FOXO transcription factor DAF-16. Ectopic expression of *rpn-6* is sufficient to confer proteotoxic stress resistance and extend lifespan, indicating that *rpn-6* is a candidate to correct deficiencies in age-related protein homeostasis disorders.

In nature, food sources are largely unpredictable and insufficient. The constant pressures that limited energetic resources place on an organism have long been theorized to cause a significant life-history trade-off: the absolute need for repairing and preventing damage to the germ line, and for ensuring elimination of damage in progeny, necessarily dominates resource allocation strategies, whereas conversely little or no evolutionary pressure will be placed on the maintenance of the soma¹. Thus, ageing, post-reproductive organisms that escape predation witness the gradual deterioration of their own somatic tissues. In support of such theories, modulations of reproduction that eliminate germ cells provide effective mechanisms for extending lifespan^{2,3}, phenotypes that may be caused by heightened resource availability within the post-mitotic soma. Likewise, it has been proposed that animals undergoing dietary restriction adopt a strategy in which resources are re-allocated towards somatic maintenance, extending lifespan and prolonging reproduction until conditions for survival become more favourable⁴.

When proliferating germline cells of *C. elegans* are removed, worms live up to 60% longer than normal and seem to be resistant to a variety of environmental stressors^{5–7}. Whereas germline ablation affords an obvious protection, the downstream effectors of such protection remain ambiguous. The re-allocation of resources to the soma seems to be directed through a specific, genetically defined stress-responsive pathway. Germline removal extends lifespan by triggering an active signalling network, involving the nuclear localization and activation of DAF-16, a forkhead transcription factor (FOXO)⁸ and the major downstream effector of the DAF-2–insulin–insulin-like growth factor (IGF) signalling (IIS) pathway. However, although worms with an ablated germ line exhibit a *daf-16*-dependent extension in lifespan, longevity caused by germline ablation functions in a synergistic manner with mutations in the IIS receptor, *daf-2* (ref. 6). Additionally, in germline-ablated animals, but not *daf-2* mutant worms, *kri-1*, *daf-9* and the nuclear hormone receptor *daf-12* are also required for the constitutive nuclear localization of DAF-16 (refs 9, 10).

Notably, post-mitotic somatic cells are particularly susceptible to age-onset protein aggregation diseases. As the somatic cell ages, the accumulation of damaged proteins represents a particular challenge to the ageing cell, especially as they aggregate in inclusions and aggresomes capable of overwhelming the cellular machinery required for their degradation^{11,12}. These effects are probably compounded by age-related dysregulation of chaperones, a downregulation of degradation machinery itself, and a continually accelerating loss in general cellular homeostasis. As such, a rapid decline in the capacity of the cell to protect its proteome has been highly correlated with multiple age-related disorders¹³. This conversely indicates that long-lived somatic cells, such as those found in a germline-ablated animal, might exhibit a heightened capacity for clearing damaged proteins, and that this proteostatic capacity might contribute to the increased longevity in these mutants.

Increased proteasome activity in *glp-1(e2141)* worms

We hypothesized that a key aspect of the proteostasis network, the ubiquitin proteasome system (UPS), might be altered in the soma of germline-ablated animals. To test this hypothesis, we first examined the activity of the 26S/30S proteasome in several long-lived mutants using a fluorogenic peptide substrate specific for the chymotrypsin-like activity of the proteasome (Supplementary Fig. 1). We found that *glp-1(e2141)* mutant worms, which lack their germ line, displayed a marked increase (over sixfold) in the chymotrypsin-like proteasome activity (Fig. 1a and Supplementary Figs 1 and 2a–c). The proteasome inhibitors MG-132, lactacystin and PI-I blocked activity from extracts in both the *glp-1* mutant and the control strain (Supplementary Fig. 3), indicating that the increased peptidase activity in *glp-1(e2141)* worms was due to the proteasome. The caspase-like and trypsin-like activities were also increased in *glp-1* mutant worms (Fig. 1b, c). Additionally, we found that a *C. elegans* genetic model of dietary restriction¹⁴ also induced proteasome activity, although to a lesser extent than the *glp-1(e2141)* mutation (Supplementary Fig. 1). In contrast, neither reduced IIS signalling by mutation of *daf-2* nor reduced mitochondrial

¹Howard Hughes Medical Institute, Glenn Center for Aging Research, Molecular and Cell Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ²Razavi Newman Center for Bioinformatics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. †Present addresses: Department of Computational Medicine and Bioinformatics, University of Michigan, 100 Washtenaw Avenue, Ann Arbor, Michigan 48109-2218, USA (A.P.C.R.); Bioinformatics and Computational Biology, Genentech, 1 DNA Way, South San Francisco, California 94080, USA (G.M.).

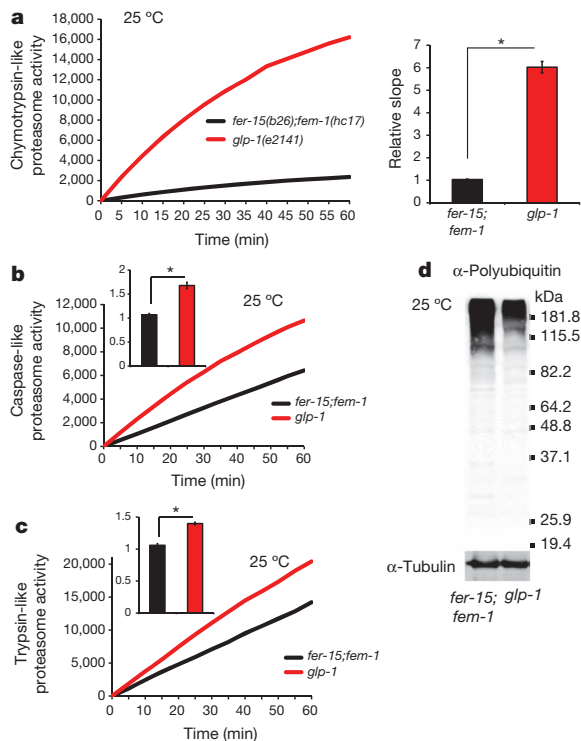


Figure 1 | Germline-lacking nematodes have increased proteasome activity. **a**, Chymotrypsin-like activity of the proteasome monitored by Z-Gly-Gly-Leu-AMC digestion in day 7 adult worm extract containing equal amounts of total protein. Proteasome activity (relative slope to control strain *fer-15(b26);fem-1(hc17)*) is shown, with error bars indicating mean \pm s.e.m. ($n = 50$, $*P = 1.83 \times 10^{-17}$). **b**, Caspase-like (Z-Leu-Leu-Glu-AMC) proteasome activity (relative slope) is shown, with error bars representing mean \pm s.e.m. ($n = 8$, $*P < 0.00001$). **c**, Trypsin-like (Ac-Arg-Leu-Arg-AMC) proteasome activity (relative slope) is shown, with error bars showing mean \pm s.e.m. ($n = 7$, $*P < 0.00001$). **d**, Representative polyubiquitinated protein immunoblot. α -Tubulin is used as the loading control.

electron transport chain activity¹⁵ upregulated proteasome activity (Supplementary Fig. 1). Consistent with increased proteasome activity in *glp-1(e2141)* worms, we observed decreased levels of polyubiquitinated proteins in these worms (Fig. 1d). To examine further UPS activity in living animals, we used a photoconvertible fluorescent UPS reporter system for live imaging and quantification of protein degradation in *C. elegans*¹⁶. This proteasome reporter consists of the photoconvertible fluorescent protein Dendra2 targeted for proteasomal degradation by fusion to a mutant form of ubiquitin (UbG76V) that cannot be cleaved by ubiquitin hydrolases. Dendra2 can be irreversibly photoconverted from a green to a red fluorescent state, providing quantification of UPS activity independently of protein synthesis. We found that this reporter is degraded more rapidly in *glp-1* mutant worms compared to control strains, whereas Dendra2 lacking the UbG76V signal remained stable (Supplementary Fig. 4).

glp-1(e2141) worms differ from other types of reproductive mutants in that their entire germ line is missing. Notably, *glp-1* mutants exhibit a significantly increased lifespan in comparison to worms that are also sterile but which still contain a proliferating germ line (Supplementary Fig. 5). We observed that the increased proteasome activity of *glp-1* mutants was not due to a benefit of sterility per se, because the normal-lived, sterile control *fer-15(b26);fem-1(hc17)* animals have similar proteasome activity as wild-type animals (Supplementary Figs 2a and 6). Furthermore, treatment with 5-fluoro-2'-deoxyuridine (FUDR), a drug used to block progeny production in worms¹⁷, did not affect proteasome activity (Supplementary Fig. 6). The *glp-1(e2141)* allele is temperature-sensitive for reproduction and longevity¹⁸; these worms are only long-lived when they are shifted to restrictive temperature

(25 °C) either during development or in early adulthood⁵. Accordingly, proteasome activity is not increased in *glp-1(e2141)* worms grown continuously at the permissive temperature (Supplementary Fig. 7). However, resembling the longevity phenotype⁵, *glp-1(e2141)* worms maintained high proteasome activity when down-shifted to a permissive temperature after germline removal (Supplementary Fig. 8). These results indicate that different forms of sterility do not have similar effects on proteasome activity, but are specific to loss of the germ line.

DAF-16 regulates proteasome activity

Because DAF-16, the worm FOXO transcription factor, is essential for the increased longevity of *glp-1* mutant worms⁵, we tested whether *daf-16* was also required for the increased proteasome activity found in *glp-1(e2141)* animals. Proteasome activity of *glp-1* mutant animals was suppressed to wild-type levels in *daf-16;glp-1* double mutant animals (Fig. 2a–c). *daf-16* is required during reproductive adulthood to modulate the ageing process in worms¹⁹. Accordingly, *daf-16* RNAi treatment of *glp-1(e2141)* animals during adulthood decreased proteasome activity (Fig. 2d). In contrast, *daf-16* RNAi did not affect proteasome activity in control strains (Fig. 2d and Supplementary Fig. 9), where DAF-16 is located in the cytosol and is inactive. Loss of *daf-16* suppressed longevity and proteasome activity, but not the reproductive phenotype of *glp-1(e2141)* worms, providing further evidence that increased proteasomal activity could not be separated from the increased longevity mediated by DAF-16 in *glp-1* mutants. We examined whether other genes required to promote DAF-16/FOXO nuclear localization in the germline longevity pathway (that is, *daf-12*, *daf-9* and *kri-1*) were also necessary for increased proteasome activity. Accordingly, reduction of any one of these genes in *glp-1(e2141)* worms resulted in decreased proteasome activity, although not to the extent of *daf-16* reduction (Fig. 2e). Furthermore, reduction of *daf-12*, *daf-9* or *kri-1* did not further decrease proteasome activity of the *daf-16;glp-1* double mutant animals or the control strain (Supplementary Fig. 10a, b). In addition to *daf-16*, *glp-1*-mediated longevity requires two additional transcription factors: *hsf-1* (ref. 20) and *skn-1* (Supplementary Fig. 11). *hsf-1* is required for the regulation of adult lifespan, heat-shock and proteotoxic stress^{21,22}; *skn-1* is the worm orthologue of *nrf-2* and has a central role in oxidative stress responses in worms, flies and mice^{23,24}. In contrast to *daf-16*, *daf-12*, *daf-9* and *kri-1*, we observed that neither *hsf-1* nor *skn-1* was required for the increased proteasome activity in *glp-1(e2141)* worms (Fig. 2f, g). To uncover a possible redundancy of either *hsf-1* or *skn-1* with *daf-16*, we knocked down these factors in the *daf-16;glp-1* double mutant animals. However, in the context of *daf-16* loss, neither *hsf-1* nor *skn-1* further affected proteasome activity in *glp-1(e2141)* worms (Supplementary Fig. 12). The nuclear hormone receptor *nhr-80* links fatty acid desaturation to lifespan extension through germline ablation in a *daf-16*-independent manner²⁵. Consistent with a requirement for *daf-16* in proteasome activity, *nhr-80* was not required for increased proteasome activity in *glp-1* mutant worms (Supplementary Fig. 13). Taken together, alterations that specifically affect DAF-16 activity, but not HSF-1, SKN-1 or NHR-80, alter proteasome activity in *glp-1* mutants, indicating that a major output for DAF-16-mediated longevity in this mutant is to increase proteasome activity.

DAF-16 regulates *rpn-6.1* levels

The 26S/30S proteasome consists of a 20S core structure that contains the proteolytic active sites and 19S cap structures that impart regulation on the activity of the holo-complex (26S, single capped, and 30S, double capped)²⁶. Although 20S particles can exist in a free form, 20S particles in their most physiological form are inactive, unable to degrade denatured proteins or cleave peptides²⁷. The 19S regulatory subunit is responsible for stimulating the 20S proteasome to degrade proteins, as ATPases of the regulatory particle open the 20S core, allowing substrates access to proteolytic active sites²⁸. Analysis of

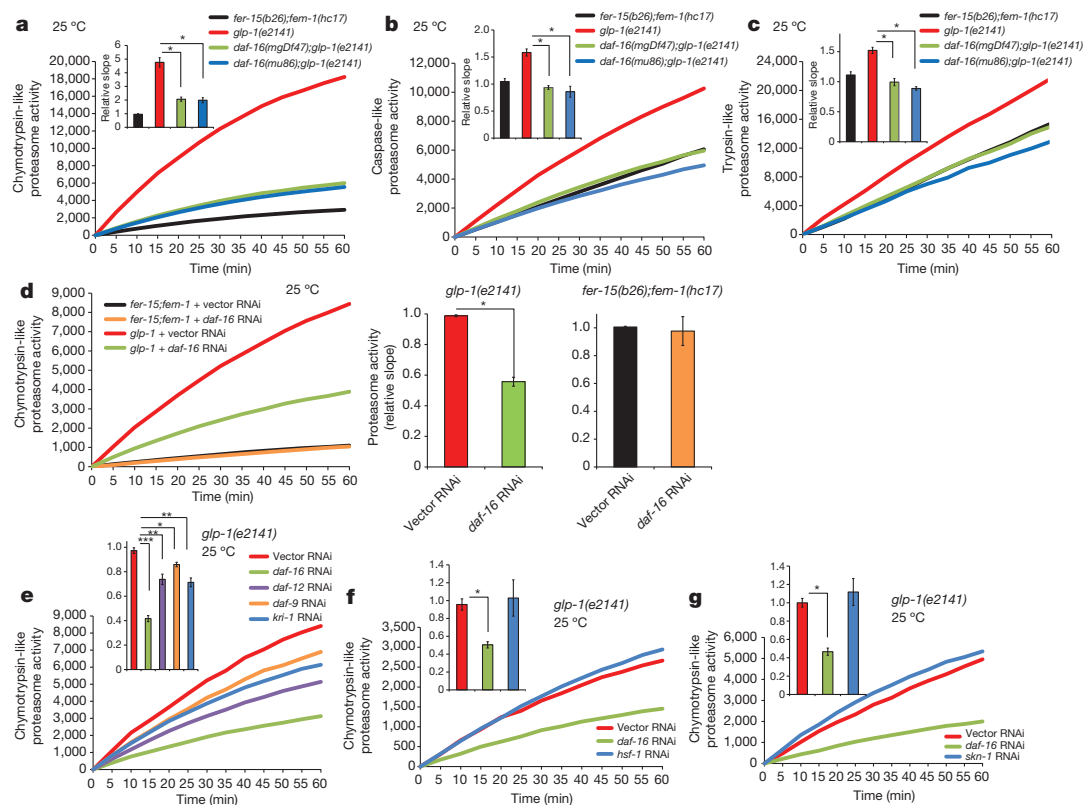


Figure 2 | DAF-16 is required for proteasome activity in *glp-1(e2141)* mutant nematodes. **a**, Chymotrypsin-like proteasome activity (relative slope to *fer-15(b26);fem-1(hc17)*) is shown. Error bars indicate mean \pm s.e.m. ($n = 11$, $*P < 0.00005$). **b**, Caspase-like proteasome activity (relative slope to *fer-15(b26);fem-1(hc17)*) is shown. Error bars indicate mean \pm s.e.m. ($n = 6$, $*P < 0.0001$). **c**, Trypsin-like proteasome activity (relative slope to *fer-15(b26);fem-1(hc17)*) is shown. Error bars represent mean \pm s.e.m. ($n = 6$, $*P < 0.005$). **d**, *glp-1(e2141)* worms fed *daf-16* RNAi bacteria show decreased chymotrypsin-like proteasome activity ($P = 8.5 \times 10^{-8}$ *glp-1* mutant fed vector RNAi bacteria versus *glp-1* mutant fed *daf-16* RNAi bacteria; $n = 16$). *daf-16* RNAi knock down does not affect proteasome activity in *fer-15;fem-1* worms ($P = 0.79$ *fer-15;fem-1* fed vector RNAi bacteria versus *fer-15;fem-1* fed *daf-16* RNAi bacteria; $n = 9$). **e**, Chymotrypsin-like proteasome activity in *glp-1(e2141)* worms fed *daf-16*, *daf-12*, *daf-9* or *kri-1* RNAi bacteria ($n = 4$; vector

RNAi versus *daf-16* RNAi ($P = 4.17 \times 10^{-5}$); vector RNAi versus *daf-12* RNAi ($P < 0.01$); vector RNAi versus *daf-9* RNAi ($P < 0.05$); vector RNAi versus *kri-1* RNAi ($P < 0.01$)). **f**, Chymotrypsin-like proteasome activity (relative slope to *glp-1* mutant fed vector RNAi bacteria) is shown. Error bars indicate mean \pm s.e.m. ($n = 8$; vector RNAi versus *daf-16* RNAi ($P < 0.0001$); vector RNAi versus *hsf-1* RNAi ($P = 0.74$)). **g**, *skn-1* RNAi does not affect chymotrypsin-like proteasome activity of *glp-1(e2141)* worms (vector RNAi versus *daf-16* RNAi ($P < 0.00001$); vector RNAi versus *hsf-1* RNAi ($P = 0.46$)). Chymotrypsin-like proteasome activity (relative slope to *glp-1* mutant fed vector RNAi bacteria) is shown. Error bars indicate the mean \pm s.e.m. ($n = 9$; vector RNAi versus *daf-16* RNAi ($P < 0.00001$); vector RNAi versus *hsf-1* RNAi ($P = 0.46$)). All activities measured at day 5 of adulthood. All RNAi treatment was initiated at day 1 of adulthood. All statistical comparisons were made by Student's *t*-test for unpaired samples.

the messenger RNA levels of the 20S proteasome subunits revealed that α -subunits were not increased in *glp-1* mutants whereas only one of the β -subunits, *pbs-5*, was moderately increased (Supplementary Fig. 14 and Supplementary Table 1). *PBS-5* is the β -type subunit that contains the chymotrypsin-like proteolytic active site²⁶. With regard to the 19S proteasome subunits, we did not detect an increase of the ATPase subunits (Fig. 3a and Supplementary Table 1). Notably, only one of the non-ATPase subunits showed increased mRNA levels in *glp-1* mutant animals: *rpn-6.1*, an essential subunit for the activity of the 26S/30S proteasome that stabilizes the otherwise weak interaction between the 20S core and the 19S cap^{29,30} (Fig. 3a, b and Supplementary Table 1). mRNA levels of the closely related *rpn-6.2* were not increased in *glp-1* mutants (Fig. 3a and Supplementary Table 1). *rpn-6.1* had a threefold increase in its expression in *glp-1* mutants and was by far the most increased of all subunits. Accordingly, knock down of *rpn-6.1* markedly decreased proteasome activity in *glp-1(e2141)* animals (Fig. 3c) similar to loss of *daf-16*. In contrast, loss of other non-ATPase subunits did not affect proteasome activity of *glp-1* mutants (Supplementary Fig. 15). Knock down of *rpn-6.1* induced an upregulation in the expression of the rest of the 26S proteasome subunits, probably to compensate for the reduction in proteasome activity induced by decreased levels of this critical subunit (Supplementary Table 2). Moreover, overexpression of *rpn-6.1* in wild-type animals

was sufficient to increase proteasome activity (Fig. 3d). The increase in *rpn-6.1* mRNA levels did not alter the expression of other 26S proteasome subunits (Supplementary Table 3). Taken together, RPN-6.1 seems to be a key component required for activation of the proteasome machinery of the germline-lacking nematodes.

We found that DAF-16 is necessary for the increased expression of *rpn-6.1* by analysing its mRNA levels in both *daf-16;glp-1* double mutants and *daf-16* RNAi-treated animals (Fig. 3e, Supplementary Figs 16 and 17 and Supplementary Table 4). Notably, mRNA levels of other proteasome subunits, including *pbs-5*, were not decreased by loss of *daf-16* in *glp-1* mutant worms (Supplementary Fig. 17). *daf-16* RNAi did not change *rpn-6.1* expression in control worms (Supplementary Fig. 16). Therefore, these results display a correlation between *daf-16* activity, *rpn-6.1* levels and proteasome activity. Consistent with DAF-16 regulating the expression of *rpn-6.1*, we identified a potential DAF-16 binding site³¹ within the first intron of *rpn-6.1* (Supplementary Fig. 18). This site is supported by a DAF-16 binding region defined by the modENCODE project³², indicating that *rpn-6.1* is likely to be a direct DAF-16 target. To explore further *rpn-6.1* transcriptional regulation, we generated a transcriptional reporter construct. We found that *rpn-6.1* was expressed in the pharynx and posterior intestine in control worms. Notably, *rpn-6.1* expression increased markedly in the pharynx and throughout the intestine of *glp-1* mutants. In

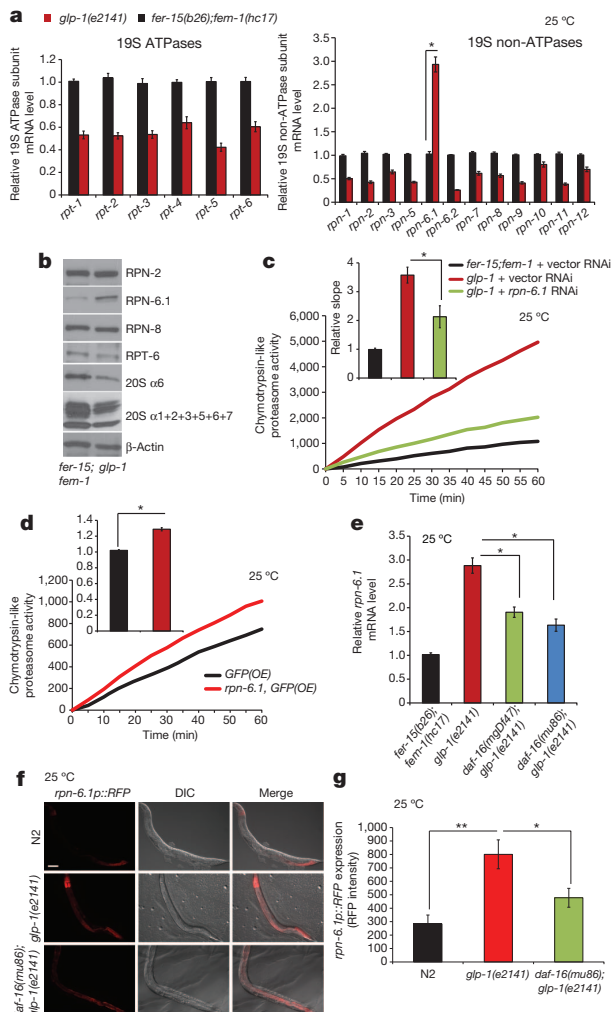


Figure 3 | DAF-16 is necessary for increased expression of *rpn-6.1* in *glp-1* mutants. **a**, Data represent the mean \pm s.e.m. of the relative expression levels to *fer-15(b26);ferm-1(hc17)* ($n = 11$, $P = 2.09 \times 10^{-6}$). **b**, Western blot analysis of RPN-2, RPN-6.1, RPN-8, RPT-6, $\alpha 6$ and $\alpha 1 + 2 + 3 + 5 + 6 + 7$. β -Actin is the loading control. **c**, *glp-1* mutants fed *rpn-6.1* RNAi bacteria starting day 1 of adulthood have decreased chymotrypsin-like proteasome activity ($P < 0.05$). Proteasome activity (relative slope to *fer-15(b26);ferm-1(hc17)*) represents the mean \pm s.e.m. ($n = 6$). **d**, Increased chymotrypsin-like proteasome activity in *rpn-6.1* overexpressing N2 worms (*rpn-6.1*, GFP(OE)) in day 1 adult worm extract ($P < 0.001$). Proteasome activity (relative slope to GFP(OE) worms) represents the mean \pm s.e.m. ($n = 9$). **e**, *daf-16;glp-1* double mutants have decreased *rpn-6.1* mRNA levels ($P < 0.001$). Graph represents the mean \pm s.e.m. ($n = 6$). **f**, Representative images of RFP expressed under control of the *rpn-6.1* promoter. *rpn-6.1* is expressed in the pharynx and posterior intestine in N2 worms. Increased *rpn-6.1* expression in *glp-1(e2141)* animals relative to N2. *daf-16;glp-1* mutant worms have decreased *rpn-6.1* expression compared to *glp-1* mutants. DIC, differential interference contrast microscopy. Scale bar, 100 μ m. **g**, Quantification of RFP signal intensity (mean \pm s.e.m. ($n = 5$)). *glp-1(e2141)* worms have increased *rpn-6.1* expression compared to N2 ($P < 0.01$) and *daf-16(mu86);glp-1(e2141)* double mutants ($P < 0.05$). All statistical comparisons were made by Student's *t*-test for unpaired samples.

daf-16;glp-1 double mutants, we found almost a twofold decreased expression of *rpn-6.1* compared to *glp-1(e2141)* worms, although *rpn-6.1* expression is still increased compared to wild-type worms (Fig. 3f, g). These results correlated with quantitative real-time polymerase chain reaction (qPCR) data indicating that *daf-16* mutations decrease *rpn-6.1* expression in *glp-1* mutants, but not to control levels (Fig. 3e, g), suggesting that a potential additional factor may be necessary for *rpn-6.1* expression in addition to DAF-16.

rpn-6.1 determines stress resistance

With the strong connection between *daf-16*, a key ageing modulator, *rpn-6.1*, a key proteasomal factor, and increased proteasome activity in long-lived *glp-1* mutant animals dependent upon both *daf-16* and *rpn-6.1*, we asked what role, if any, does *rpn-6.1* have in longevity. To assess the requirement for *rpn-6.1* during lifespan, we conducted RNAi knock down of this gene in *C. elegans*. Because proteasomal function is required during larval development³³, we initiated *rpn-6.1* RNAi treatment during adulthood, the time at which *daf-16* is required for longevity assurance¹⁹. Knock down of *rpn-6.1* substantially decreased the lifespan of *glp-1* mutant animals (Supplementary Fig. 19). However, loss of *rpn-6.1* also decreased lifespan of both wild-type and sterile control animals (*fer-15(b26);ferm-1(hc17)*). In addition to germ-cell loss, we also examined the effects of *rpn-6.1* RNAi on other pathways that influence lifespan such as reduced IIS (*daf-2(e1370)* worms), reduced mitochondrial electron transport chain (*isp-1(qm150)* worms) and reduced food intake (*eat-2(ad1116)* worms). In all cases, knock down of *rpn-6.1* substantially decreased lifespan, confirming that this gene is essential for the viability of adult animals, making lifespan analysis by *rpn-6.1* loss of function difficult to interpret (Supplementary Fig. 19).

To explore whether *rpn-6.1* might have a positive role in longevity, we tested the impact of increased *rpn-6.1* expression. We overexpressed *rpn-6.1* in wild-type worms and conducted a series of physiological assays to measure the effects of *rpn-6.1* overexpression (OE) on resistance to challenges of oxidative stress, heat-shock and ultraviolet (UV) damage, all of which are correlated with increased longevity. *rpn-6.1*(OE) nematodes were significantly more resistant, than control strains, to oxidative stress induced by growing the worms in the presence of paraquat (Fig. 4a and Supplementary Fig. 20a). Under heat stress (34 °C), *rpn-6.1*(OE) worms lived markedly longer than control strains (Fig. 4b and Supplementary Fig. 20b). Interestingly, increased levels of *rpn-6.1* did not result in global upregulation of all stress responses because *rpn-6.1*(OE) did not protect against UV damage (Fig. 4c and Supplementary Fig. 20c). Because overexpression of *rpn-6.1* increased resistance to conditions that challenge the proteome, we examined whether it also resulted in lifespan extension. *rpn-6.1*(OE) did not extend the lifespan of worms at 20 °C but it did at 25 °C, a temperature that results in mild heat stress (Fig. 4d and Supplementary Fig. 20d). *daf-16* RNAi treatment blocked the lifespan extension induced by *rpn-6.1*(OE) at 25 °C (Fig. 4e and Supplementary Fig. 20e). Therefore, under conditions of proteome stress, overexpression of *rpn-6.1* is sufficient to promote increased survival. As a more formal test, we asked whether animals with a reduced heat-shock response via *hsf-1* downregulation had increased survival when *rpn-6.1* was overexpressed. *hsf-1*-RNAi-treated *rpn-6.1*(OE) worms were long-lived compared to control strains under the same treatment (Fig. 4f and Supplementary Fig. 20f). This last result not only indicates that the lifespan extension induced by *rpn-6.1*(OE) is *hsf-1* independent, but also suggests that these worms can significantly overcome the loss of this critical transcription factor required for adult lifespan, heat-shock and proteotoxicity responses.

Intrigued by the protection that *rpn-6.1* overexpression could confer, we hypothesized that RPN-6.1 could be a potential candidate to correct protein homeostasis deficiencies underlying diseases such as Alzheimer's, Parkinson's or Huntington's disease. Because the latter disease has been associated with proteasome failure³⁴, we tested whether increased levels of *rpn-6.1* could have beneficial effects in a polyglutamine (polyQ) disease model. Worm motility is markedly reduced by the aggregation of polyQ expression in neurons, with a pathogenic threshold at a length of 35–40 glutamines³⁵. Notably, *rpn-6.1* overexpression substantially improved motility and reduced toxicity of worms expressing polyQ40 and polyQ67 (Fig. 5a and Supplementary Fig. 21). In addition, loss of *rpn-6.1* had a detrimental effect on the motility phenotype of polyQ67 worms even at early (day 3) adulthood stages (Fig. 5b and Supplementary Fig. 22). Furthermore,

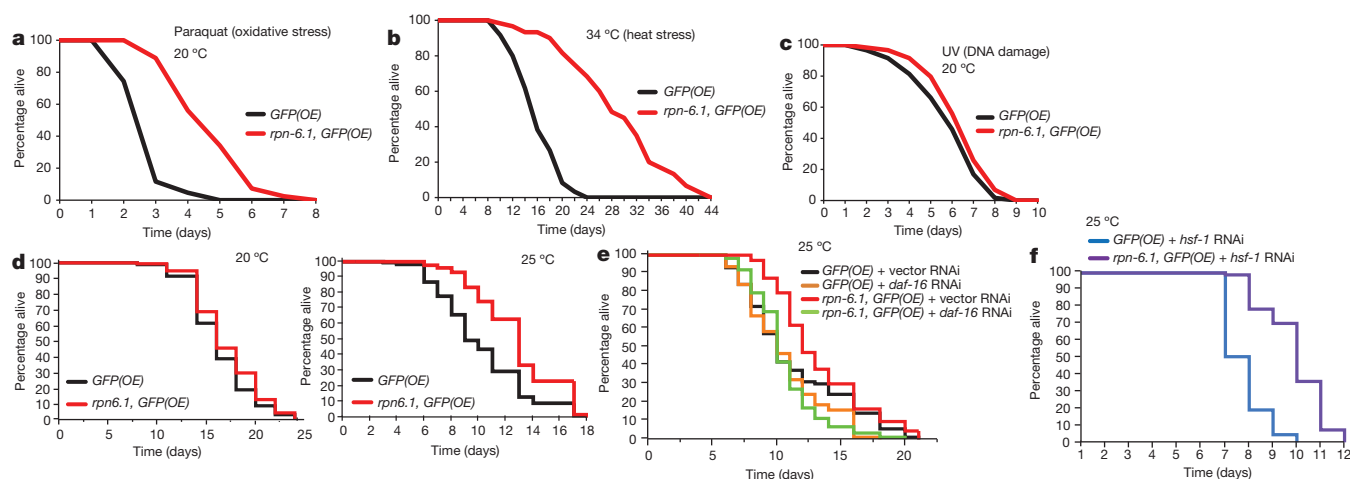


Figure 4 | *rpn-6.1* is a determinant of stress resistance and viability. **a**, *rpn-6.1* overexpressing (*OE*) worms live longer than controls under oxidative stress (log rank, $P < 0.0001$, *GFP(OE)*: mean = 2.97 ± 0.10 , $n = 43/60$; *rpn-6.1, GFP(OE)*: mean = 4.89 ± 0.17 , $n = 40/59$). **b**, *rpn-6.1(OE)* worms live longer than controls under heat stress conditions (34 °C) (log rank, $P < 0.0001$, *GFP(OE)*: mean = 16.20 ± 0.47 , $n = 60/60$; *rpn-6.1, GFP(OE)*: mean = 29.83 ± 0.99 , $n = 60/60$). **c**, Ultraviolet stress assay on *rpn-6.1(OE)* worms (log rank, $P = 0.06$, *GFP(OE)*: mean = 6.01 ± 0.27 , $n = 59/60$; *rpn-6.1, GFP(OE)*: mean = 6.57 ± 0.19 , $n = 58/59$). **d**, *rpn-6.1(OE)* does not affect lifespan at 20 °C (log rank, $P = 0.15$, *GFP(OE)*: mean = 16.33 ± 0.34 , $n = 90/100$; *rpn-6.1, GFP(OE)*: mean = 17.05 ± 0.34 , $n = 95/100$). *rpn-6.1* overexpression extends lifespan at 25 °C (log rank, $P < 0.0001$, *GFP(OE)*: mean = 10.19 ± 0.32 , $n = 109/120$; *rpn-6.1, GFP(OE)*: mean = 12.73 ± 0.29 ,

$n = 113/120$). **e**, *daf-16* RNAi (starting at day 1 of adulthood) blocks lifespan extension induced by *rpn-6.1* overexpression (*rpn-6.1, GFP(OE)* fed vector RNAi bacteria versus *rpn-6.1, GFP(OE)* fed *daf-16* RNAi bacteria, log rank, $P < 0.0001$). *GFP(OE)* fed vector RNAi bacteria: mean = 11.36 ± 0.45 , $n = 86/100$; *GFP(OE)* fed *daf-16* RNAi bacteria: mean = 10.59 ± 0.36 , $n = 77/100$; *rpn-6.1, GFP(OE)* fed vector RNAi bacteria: mean = 13.27 ± 0.37 , $n = 87/100$; *rpn-6.1, GFP(OE)* fed *daf-16* RNAi bacteria: mean = 10.59 ± 0.28 , $n = 88/100$. **f**, *hsf-1*-RNAi-treated *rpn-6.1(OE)* worms were long-lived compared to controls (log rank, $P < 0.0001$). RNAi was initiated at day 1 of adulthood. *GFP(OE)* fed *hsf-1* RNAi bacteria: mean = 7.74 ± 0.09 , $n = 93/99$; *rpn-6.1, GFP(OE)* fed *hsf-1* RNAi bacteria: mean = 9.92 ± 0.13 , $n = 94/100$. See Supplementary Table 5 for statistical analysis and replicate data of stress assays and lifespan experiments.

we observed by filter trap analysis that *rpn-6.1(OE)* reduced polyQ aggregate levels whereas polyQ67 total protein levels remained constant (Fig. 5c), suggesting that *rpn-6.1* specifically reduces aggregated, but not soluble, polyQ proteins.

Discussion

A growing body of evidence suggests that the protective modulation of various nodes of the proteostasis network, including the heat-shock response and autophagy, can contribute to the extended lifespan caused by the IIS^{36,37}, diet restriction³⁸ and germline-signalling pathways³⁹. We report evidence for the requirement of an upregulated proteasome activity in the extended lifespan of germline-deficient nematodes. Our initial analysis of proteasome activity among different longevity models in the worm reveals that only *glp-1* mutant and diet-restricted animals share an increased proteasome activity, and we hypothesize that these animals may share a strategy in which resources are actively re-allocated from the germ line to the soma, resulting in an enhanced protection of the proteome within somatic cells. Furthermore, we find distinct differences in the proteasome activity between *glp-1* and *daf-2* mutant animals that is mediated by DAF-16, and in part by KRI-1, DAF-12 and DAF-9, confirming previous genetic suggestions that DAF-16 activity is differentially regulated between *glp-1* and *daf-2* mutants. Mechanistically, in germline-deficient animals, *rpn-6.1* and subsequent increases in proteasome activity seem to be direct downstream targets of DAF-16/FOXO. Our results thus provide new insights into proteostasis regulation and provide a link between the longevity regulator DAF-16 and proteasome activity regulation upon *rpn-6.1* expression.

We further define RPN-6 as a potent factor to increase resistance to proteotoxic stress, as its upregulation can delay the deleterious effects of strong adverse conditions. It is intriguing to speculate that one method to ensure survival of the soma may be the direct activation of FOXO/*daf-16*, under limited nutrient availability or loss of the germ line, resulting in increased *rpn-6.1* levels and increased proteome maintenance. Recently, it has been reported that changes in the proteasome may explain why ageing is a risk factor for neurodegenerative diseases such

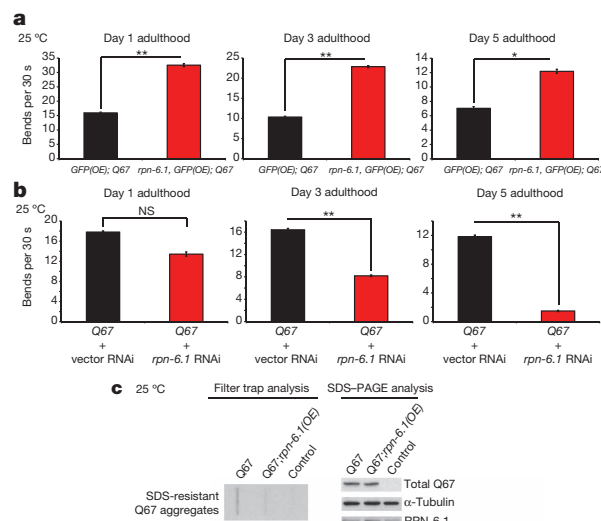


Figure 5 | *rpn-6.1* protects from polyglutamine aggregation. **a**, *rpn-6.1* improves motility in polyQ67 worms. Bar graphs represent average (\pm s.e.m.) thrashing over a 30-s period on day 1 ($P < 0.0001$, *GFP(OE); Q67* ($n = 40$), *rpn-6.1, GFP(OE); Q67* ($n = 41$)), day 3 ($P < 0.0001$, *GFP(OE); Q67* ($n = 40$), *rpn-6.1, GFP(OE); Q67* ($n = 44$)) and day 5 ($P < 0.05$, *GFP(OE); Q67* ($n = 31$), *rpn-6.1, GFP(OE); Q67* ($n = 39$)) of adulthood. **b**, Loss of *rpn-6.1* has a detrimental effect on the motility defects of polyQ67 worms. Bar graphs represent average (\pm s.e.m.) thrashing over a 30-s period on day 1 ($P = 0.15$, *Q67* fed vector RNAi bacteria ($n = 58$), *Q67* fed *rpn-6.1* RNAi bacteria ($n = 30$)), day 3 ($P < 0.0001$, *Q67* fed vector RNAi bacteria ($n = 44$), *Q67* fed *rpn-6.1* RNAi bacteria ($n = 51$)) and day 5 ($P < 0.0001$, *Q67* fed vector RNAi bacteria ($n = 41$), *Q67* fed *rpn-6.1* RNAi bacteria ($n = 47$)) of adulthood. All statistical comparisons were made by Student's *t*-test for unpaired samples. **c**, Filter trap analysis indicates that *rpn-6.1* overexpression results in reduced polyQ aggregates (detected by anti-GFP antibody). Right panel: SDS-PAGE analysis with antibodies to GFP, RPN-6 and α -tubulin loading control. * $P < 0.05$; ** $P < 0.0001$.

as Alzheimer's, Parkinson's and Huntington's disease⁴⁰. Therefore, RPN-6 may be a powerful candidate to correct deficiencies in disorders associated with a failure in protein homeostasis. It will be of crucial interest to explore in mammalian models whether RPN-6 could indeed alleviate the associated symptoms of these disorders.

METHODS SUMMARY

Caenorhabditis elegans were cultured using standard techniques⁴¹ and fed on *Escherichia coli* OP50 or HT115 containing a double-stranded-RNA-expressing plasmid⁴².

26S proteasome activity assays. *In vitro* 26S proteasome activity assays were performed as previously described²⁷. Worms were lysed in proteasome activity assay buffer (50 mM Tris-HCl, pH 7.5, 250 mM sucrose, 5 mM MgCl₂, 0.5 mM EDTA, 2 mM ATP and 1 mM dithiothreitol) using a Precellys 24 homogenizer (Bertin technologies). Lysate was centrifuged at 10,000g for 15 min at 4 °C. A total of 25 µg of total protein lysate was transferred to a 96-well microtitre plate (BD Falcon) and incubated with fluorogenic substrate. Fluorescence (380-nm excitation, 460-nm emission) was monitored on a microplate fluorometer (Infinite M1000, Tecan) every 5 min for 1 h at 25 °C.

Motility assay. Thrashing rate was determined as previously described³⁵. Worms were transferred to a drop of M9 buffer and after 30 s of adaptation the number of body bends was counted for 30 s. A body bend was defined as a change in direction of the bend at the midbody of a nematode⁴³.

Filter trap. Worm extracts were generated by glass bead disruption on ice in non-denaturing lysis buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100) supplemented with EDTA-free protease inhibitor cocktail (Roche). Lysate was centrifuged at 5,000g for 5 min. A total of 70 µg of protein extract was supplemented with SDS at a final concentration of 1% and loaded onto a cellulose acetate membrane assembled in a slot blot apparatus (BioRad). The membrane was washed with 0.1% SDS and retained Q67-GFP was assessed by immunoblotting for GFP (Roche).

A detailed description of all experimental methods including *C. elegans* strains, growth, imaging, lifespan analysis, stress assays and RNAi application is provided in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 8 August 2011; accepted 13 June 2012.

Published online 26 August 2012.

- Kirkwood, T. B. Evolution of ageing. *Nature* **270**, 301–304 (1977).
- Kenyon, C. A pathway that links reproductive status to lifespan in *Caenorhabditis elegans*. *Ann. NY Acad. Sci.* **1204**, 156–162 (2010).
- Partridge, L., Gems, D. & Withers, D. J. Sex and death: what is the connection? *Cell* **120**, 461–472 (2005).
- Shanley, D. P. & Kirkwood, T. B. Calorie restriction and aging: a life-history analysis. *Evolution* **54**, 740–750 (2000).
- Arantes-Oliveira, N., Apfeld, J., Dillin, A. & Kenyon, C. Regulation of life-span by germ-line stem cells in *Caenorhabditis elegans*. *Science* **295**, 502–505 (2002).
- Hsin, H. & Kenyon, C. Signals from the reproductive system regulate the lifespan of *C. elegans*. *Nature* **399**, 362–366 (1999).
- Wang, M. C., O'Rourke, E. J. & Ruvkun, G. Fat metabolism links germline stem cells and longevity in *C. elegans*. *Science* **322**, 957–960 (2008).
- Lin, K., Hsin, H., Libina, N. & Kenyon, C. Regulation of the *Caenorhabditis elegans* longevity protein DAF-16 by insulin/IGF-1 and germline signaling. *Nature Genet.* **28**, 139–145 (2001).
- Berman, J. R. & Kenyon, C. Germ-cell loss extends *C. elegans* life span through regulation of DAF-16 by *kri-1* and lipophilic-hormone signaling. *Cell* **124**, 1055–1068 (2006).
- Gerisch, B., Weitzel, C., Kober-Eisermann, C., Rottiers, V. & Antebi, A. A hormonal signaling pathway influencing *C. elegans* metabolism, reproductive development, and life span. *Dev. Cell* **1**, 841–851 (2001).
- Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin-proteasome system by protein aggregation. *Science* **292**, 1552–1555 (2001).
- Bennett, E. J., Bence, N. F., Jayakumar, R. & Kopito, R. R. Global impairment of the ubiquitin-proteasome system by nuclear or cytoplasmic protein aggregates precedes inclusion body formation. *Mol. Cell* **17**, 351–365 (2005).
- Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W. & Balch, W. E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **78**, 959–991 (2009).
- Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **95**, 13091–13096 (1998).
- Dillin, A. *et al.* Rates of behavior and aging specified by mitochondrial function during development. *Science* **298**, 2398–2401 (2002).
- Hamer, G., Matilainen, O. & Holmberg, C. I. A photoconvertible reporter of the ubiquitin-proteasome system *in vivo*. *Nature Methods* **7**, 473–478 (2010).
- Mitchell, D. H., Stiles, J. W., Santelli, J. & Sanadi, D. R. Synchronous growth and aging of *Caenorhabditis elegans* in the presence of fluorodeoxyuridine. *J. Gerontol.* **34**, 28–36 (1979).
- Priess, J. R., Schnabel, H. & Schnabel, R. The *glp-1* locus and cellular interactions in early *C. elegans* embryos. *Cell* **51**, 601–611 (1987).
- Dillin, A., Crawford, D. K. & Kenyon, C. Timing requirements for insulin/IGF-1 signaling in *C. elegans*. *Science* **298**, 830–834 (2002).
- Hansen, M., Hsu, A. L., Dillin, A. & Kenyon, C. New genes tied to endocrine, metabolic, and dietary regulation of lifespan from a *Caenorhabditis elegans* genomic RNAi screen. *PLoS Genet.* **1**, e17 (2005).
- Cohen, E., Bieschke, J., Perciavalle, R. M., Kelly, J. W. & Dillin, A. Opposing activities protect against age-onset proteotoxicity. *Science* **313**, 1604–1610 (2006).
- Hsu, A. L., Murphy, C. T. & Kenyon, C. Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* **300**, 1142–1145 (2003).
- Alam, J. *et al.* Nrf2, a Cap'n'Collar transcription factor, regulates induction of the heme oxygenase-1 gene. *J. Biol. Chem.* **274**, 26071–26078 (1999).
- An, J. H. & Blackwell, T. K. SKN-1 links *C. elegans* mesodermal specification to a conserved oxidative stress response. *Genes Dev.* **17**, 1882–1893 (2003).
- Goudeau, J. *et al.* Fatty acid desaturation links germ cell loss to longevity through NHR-80/HNF4 in *C. elegans*. *PLoS Biol.* **9**, e1000599 (2011).
- Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
- Kisselev, A. F. & Goldberg, A. L. Monitoring activity and inhibition of 26S proteasomes with fluorogenic peptide substrates. *Methods Enzymol.* **398**, 364–378 (2005).
- Köhler, A. *et al.* The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Mol. Cell* **7**, 1143–1152 (2001).
- Pathare, G. R. *et al.* The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. *Proc. Natl Acad. Sci. USA* **109**, 149–154 (2012).
- Santamaria, P. G., Finley, D., Ballesta, J. P. & Remacha, M. Rpn6p, a proteasome subunit from *Saccharomyces cerevisiae*, is essential for the assembly and activity of the 26 S proteasome. *J. Biol. Chem.* **278**, 6687–6695 (2003).
- Furuyama, T., Nakazawa, T., Nakano, I. & Mori, N. Identification of the differential distribution patterns of mRNAs and consensus binding sequences for mouse DAF-16 homologues. *Biochem. J.* **349**, 629–634 (2000).
- Celniker, S. E. *et al.* Unlocking the secrets of the genome. *Nature* **459**, 927–930 (2009).
- Ghazi, A., Henis-Korenblit, S. & Kenyon, C. Regulation of *Caenorhabditis elegans* lifespan by a proteasomal E3 ligase complex. *Proc. Natl Acad. Sci. USA* **104**, 5947–5952 (2007).
- Li, X. J. & Li, S. Proteasomal dysfunction in aging and Huntington disease. *Neurobiol. Dis.* **43**, 4–8 (2011).
- Brignull, H. R., Moore, F. E., Tang, S. J. & Morimoto, R. I. Polyglutamine proteins at the pathogenic threshold display neuron-specific aggregation in a pan-neuronal *Caenorhabditis elegans* model. *J. Neurosci.* **26**, 7597–7606 (2006).
- Melendez, A. *et al.* Autophagy genes are essential for dauer development and life-span extension in *C. elegans*. *Science* **301**, 1387–1391 (2003).
- Morley, J. F. & Morimoto, R. I. Regulation of longevity in *Caenorhabditis elegans* by heat shock factor and molecular chaperones. *Mol. Biol. Cell* **15**, 657–664 (2004).
- Hansen, M. *et al.* A role for autophagy in the extension of lifespan by dietary restriction in *C. elegans*. *PLoS Genet.* **4**, e24 (2008).
- Lapierre, L. R., Melendez, A. & Hansen, M. Autophagy links lipid metabolism to longevity in *C. elegans*. *Autophagy* **8**, 144–146 (2012).
- Zabel, C. *et al.* Proteasome and oxidative phosphorylation changes may explain why aging is a risk factor for neurodegenerative disorders. *J. Proteomics* **73**, 2230–2238 (2010).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Chai, Y., Shao, J., Miller, V. M., Williams, A. & Paulson, H. L. Live-cell imaging reveals divergent intracellular dynamics of polyglutamine disease proteins and supports a sequestration model of pathogenesis. *Proc. Natl Acad. Sci. USA* **99**, 9310–9315 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Panowski for help with the generation of transgenic strains. We thank D. Joyce for proteasome activity assays and S. Wolf for comments on the manuscript. This work was supported by HHMI and the NIA. D.V. was a recipient of the F.M. Kirby, Inc. Foundation Postdoctoral Scholar Award and Beatriz de Pinós (AGAUR) fellowship.

Author Contributions D.V. and A.D. planned and supervised the project. D.V. performed the experiments, data analysis and interpretation. I.M. performed biochemistry experiments and contributed to other assays. Z.L. performed UPS reporter experiment, lifespans and injections. P.M.D. performed the filter trap assay. C.M. performed immunoblots. A.P.C.R. and G.M. performed the transcription factor binding site analysis. The manuscript was written by D.V. and A.D. and edited by I.M. and C.M. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.D. (dillin@salk.edu).

METHODS

Caenorhabditis elegans strains and generation of transgenic lines. CF512 (*fer-15(b26)II;fem-1(hc17)IV*), CB4037 (*glp-1(e2141)III*), AU147 (*daf-16(mgDf47)I;glp-1(e2141)III*), CF1880 (*daf-16(mu86)I;glp-1(e2141)III*), DA1116 (*eat-2(ad1116)II*) and wild-type (N2) *C. elegans* strains were obtained from the *Caenorhabditis* Genetic Center. AGD151 (*eat-2(ad1116)II; fer-15(b26)II;fem-1(hc17)IV*) was generated by crossing CF512 with DA1116 (*eat-2(ad1116)II*). CF596 (*daf-2(mu150)III; fer-15(b26)II;fem-1(hc17)IV*) was a gift from C. Kenyon. *C. elegans* were handled using standard methods⁴¹.

For the generation of worm strains AGD597–AGD598 (N2, *uthEx556[psur5::rpn-6.1, pmyo3::GFP]* and N2, *uthEx556[psur5::rpn-6.1, pmyo3::GFP]*), a DNA plasmid mixture containing 75 ng μl^{-1} pDV1 (*psur5::rpn-6.1*) and 20 ng μl^{-1} pPD93_97 (*pmyo-3::GFP*) was injected into the gonads of adult N2 hermaphrodite animals, using standard methods⁴⁴. GFP-positive F₁ progeny were selected. Individual F₂ worms were isolated to establish independent lines. Control worms (AGD614) used in experiments with AGD597–AGD598 were generated by microinjecting N2 worms with 20 ng μl^{-1} pPD93_97 (*pmyo-3::GFP*). AGD886 (*fer-15(b26)II; fem-1(hc17)IV; uthEx557[psur5::rpn-6.1, pmyo3::GFP]*) was generated by crossing AGD598 to CF512. Control strain AGD885 (*fer-15(b26)II;fem-1(hc17)IV; uthEx633[pmyo3p::GFP]*) was generated by crossing AGD614 to CF512.

Both AM101 (*rmls110[pF25B3.3::Q40::YFP]*) and AM716 (*rmls284[pF25B3.3::Q67::YFP]*) were a gift from R. I. Morimoto. For the generation of worm strains AGD850 (*rmls110[pF25B3.3::Q40::YFP];uthEx557[psur5::rpn-6.1,pmyo3::GFP]*) and AGD851 (*rmls284[pF25B3.3::Q67::YFP];uthEx557[psur5::rpn-6.1,pmyo3::GFP]*), AGD598 strain was crossed to AM101 and AM716, respectively. Control strains AGD866 (*rmls110[pF25B3.3::Q40::YFP]; uthEx633[pmyo3::GFP]*) and AGD867 (*rmls284[pF25B3.3::Q67::YFP];uthEx633[pmyo3::GFP]*) were generated by crossing AGD614 to AM101 and AM716, respectively.

For the generation of worm strains AGD945–AGD946 (N2, *uthEx649[rpn-6p::tdTomato, pRF4(rol-6)]* and N2, *uthEx650[rpn-6p::tdTomato, pRF4(rol-6)]*), a DNA plasmid mixture containing 75 ng μl^{-1} pDV2 (*rpn-6p::tdTomato*) and 20 ng μl^{-1} pRF4(*rol-6*) was injected into the gonads of adult N2 hermaphrodite animals. Roller-phenotype-positive F₁ progeny were selected. Individual F₂ worms were isolated to establish independent lines. AGD1047 (*glp-1(e2141)III; uthEx649[rpn-6p::tdTomato, pRF4(rol-6)]*) was generated by crossing AGD945 to CB4037. AGD1048 (*daf-16(mu86)I;glp-1(e2141)III; uthEx649[rpn-6p::tdTomato, pRF4(rol-6)]*) was generated by crossing AGD945 to CF1880.

YD1 (N2, *xxEx1[Punc-54::Dendra2]*) and YD3 (N2, *xxEx3[Punc-54::UbG76V::Dendra2]*) were a gift from C. I. Holmberg. AGD1032 (*glp-1(e2141)III; xxEx1[Punc-54::Dendra2]*) was generated by crossing YD1 to CB4037. AGD1033 (*glp-1(e2141)III; xxEx3[Punc-54::UbG76V::Dendra2]*) was generated by crossing YD3 to CB4037. AGD1036 (*fer-15(b26)II;fem-1(hc17)IV; xxEx1[Punc-54::Dendra2]*) was generated by crossing YD1 to CF512. AGD1037 (*fer-15(b26)II;fem-1(hc17)IV; xxEx3[Punc-54::UbG76V::Dendra2]*) was generated by crossing YD3 to CF512.

Construction of *rpn-6.1* expression construct. To construct pDV1, the *rpn-6.1* *C. elegans* expression plasmid pPD95.77 from the Fire Lab kit was digested with SphI and XmaI to insert 3.6 kilobases of the *sur5* promoter. The resultant vector was then digested with KpnI and EcoRI to excise GFP and insert a multi-cloning site containing KpnI, NheI, NotI, XbaI and EcoRI. F57B9.10.A (*rpn-6.1*) was PCR amplified from cDNA to include 5' XmaI and 3' XbaI restriction sites then cloned into the aforementioned vector. All constructs were sequence verified.

Construction of *rpn-6.1* transcriptional reporter construct. To construct pDV2, pPD95.77 from the Fire Lab kit was digested to replace GFP with tdTomato. The promoter region and first intron of F57B9.10.A (*rpn-6.1*) was PCR amplified from N2 gDNA to include –363 to +1012 then cloned into the aforementioned vector using SalI and BamHI. The construct includes 46 nucleotides of exon 1. Construct was sequence verified.

RNAi constructs. RNAi-treated strains were fed *E. coli* (HT115) containing an empty control vector (L4440) or expressing double-stranded RNAi⁴². *daf-12*, *rpn-2*, *rpn-6.1*, *rpn-11* and *skn-1* RNAi constructs used were taken from the Vidal RNAi library. *cco-1*, *rpn-1*, *nhr-80*, *daf-9*, *hsf-1* and *kri-1* RNAi constructs used were from the Ahringer RNAi library. pAD43, the *daf-16* RNAi construct, was previously described¹⁹. See Supplementary Table 6 for further details about double-stranded RNAi constructs used for knockdown assays.

Lifespan studies. Lifespan analyses were performed as described previously¹⁹. Worms were synchronized by egg laying during 2 h. Animals were grown at 20 °C until day 1 of adulthood. One-hundred animals were used per condition and scored every day or every other day. Lifespans were conducted at either 20 °C or 25 °C as stated in the figure legends. For non-integrated lines AGD597, AGD598 and AGD886, GFP-positive worms were selected for lifespan studies. JMP IN 8 software was used for statistical analysis to determine means and percentiles. In

all cases, *P* values were calculated using the log-rank (Mantel–Cox) method. See Supplementary Table 5 for statistical analysis and replicate data.

Stress assays. For heat-shock assays, eggs were transferred to plates seeded with *E. coli* (OP50) bacteria and grown to day 1 of adulthood at 20 °C. Worms were then transferred to fresh plates and heat shocked at 34 °C. Worms were checked every hour for viability. Paraquat assays were performed as previously described⁴⁵. Briefly day-1 adults were transferred to plates containing 7.5 mM paraquat and cultured at 25 °C. Worms were checked every day for viability. For UV irradiation assays⁴⁶, day-5 adult worms were transferred to plates without OP50 and exposed to 1,200 J m^{–2} of UV using a UV Stratalinker. Worms were transferred back to fresh plates seeded with *E. coli* (OP50) and scored daily for viability.

Motility assay. Thrashing rate was determined as previously described³⁵. Animals were grown at 20 °C until L4 stage and then grown at 25 °C for the rest of the experiment. Worms were fed with *E. coli* (OP50) bacteria. RNAi-treated strains were fed *E. coli* (HT115) containing an empty control vector (L4440) or expressing double-stranded RNAi of the *rpn-6.1* gene. Worms were transferred at day 1, 3 or 5 of adulthood to a drop of M9 buffer and after 30 s of adaptation the number of body bends was counted for 30 s. A body bend was defined as change in direction of the bend at the midbody of an animal⁴³.

26S proteasome fluorogenic peptidase assays. *In vitro* 26S proteasome activity assays were performed as previously described²⁷. Briefly, worms were lysed in proteasome activity assay buffer (50 mM Tris-HCl, pH 7.5, 250 mM sucrose, 5 mM MgCl₂, 0.5 mM EDTA, 2 mM ATP and 1 mM dithiothreitol) using a Precellys 24 homogenizer (Bertin technologies). Lysate was centrifuged at 10,000g for 15 min at 4 °C. For each experiment, 25 μg of total protein lysate was transferred to a 96-well microtitre plate (BD Falcon) then fluorogenic substrate was added. For measuring the chymotrypsin-like activity of the proteasome either Z-Gly-Gly-Leu-AMC (Enzo) or Suc-Leu-Leu-Val-Tyr-AMC (Enzo) was used. Z-Leu-Leu-Glu-AMC (Enzo) was used to measure the caspase-like activity of the proteasome and Ac-Arg-Leu-Arg-AMC for the proteasome trypsin-like activity. Fluorescence (380-nm excitation, 460-nm emission) was monitored on a microplate fluorometer (Infinite M1000, Tecan) every 5 min for 1 h at 25 °C.

Western blot. For each strain, 2,000 adult worms were collected in proteasome assay activity buffer supplemented with protease inhibitors (Roche) and lysed using a Precellys 24 homogenizer. Lysate was centrifuged at 10,000g for 15 min at 4 °C. 40 μg of total protein was resolved by SDS–PAGE and transferred to nitrocellulose membrane. Western blot analysis was performed with anti-20S alpha 1–7 (Abcam), anti-proteasome 20S C2 (Abcam), anti-Rpt6 (Enzo), anti-Rpt5 (Enzo), anti-PSMD7 (Abcam), anti-Rpn2 (Abcam), anti-PSMD11 (Novus), anti-FK1 (Enzo), GFP (Roche), anti- α -tubulin (Sigma) and anti- β -actin (Abcam).

Filter trap. Animals were grown at 20 °C until L4 stage and then grown at 25 °C for the rest of the experiment. Day 1 adult worms were collected with M9 buffer and worm pellets were frozen with liquid N₂. Frozen worm pellets were thawed on ice and worm extracts were generated by glass bead disruption on ice in non-denaturing lysis buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100) supplemented with EDTA-free protease inhibitor cocktail (Roche). Worm and cellular debris were removed with 5,000g spin for 5 min. Approximately 70 μg of protein extract was supplemented with SDS at a final concentration of 1% and loaded onto a cellulose acetate membrane assembled in a slot blot apparatus (BioRad). The membrane was washed with 0.1% SDS and retained Q67-GFP was assessed by immunoblotting for GFP (Roche). Extracts were also analysed by SDS–PAGE to determine protein expression levels.

Microscopy, image analysis, equipment and settings. Newly hatched larvae were grown at 25 °C until day 3 of adulthood. These young adults were mounted at room temperature (20–23 °C) on a 10% agarose pad on glass slides with 1 μl of M9, covered with cover slip. For imaging, Zeiss Axiovert microscope and AxioCam with software AxioVision Rel. 4.7 was used. Images of whole worms were acquired with 10 \times 0.45 numerical aperture (NA) plan-apochromat objectives. Photoconversion was carried out using a 405-nm filter and an EXFO X-Cite 120Q metal halide lamp with 100% output for 60 s. Worms were imaged before and after photoconversion, and then were recovered on feeding plates at 20 °C. After 24 h, photoconverted worms were imaged with the same setting. Fluorescence intensities were analysed with AxioVision Rel. 4.7.

RNA isolation and quantitative RT–PCR. Total RNA was isolated from synchronized populations of approximately 2,000 day-5 adults. Total RNA was extracted using TRIzol reagent (GIBCO). cDNA was generated using Quantitect Reverse Transcriptase kit (Qiagen). SybrGreen real-time qPCR experiments were performed with a 1:20 dilution of cDNA using an ABI Prism7900HT (Applied Biosystems) following the manufacturer's instructions.

Data were analysed with the comparative $2\Delta\Delta C_t$ method using the geometric mean of *cdc-42*, *pmp-3* and *Y45F10D.4* as endogenous control⁴⁷. See Supplementary Table 7 for details about the primers used for this assay.

44. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
45. Vazquez-Manrique, R. P. *et al.* Reduction of *Caenorhabditis elegans* frataxin increases sensitivity to oxidative stress, reduces lifespan, and causes lethality in a mitochondrial complex II mutant. *FASEB J.* **20**, 172–174 (2006).
46. Wolff, S. *et al.* SMK-1, an essential regulator of DAF-16-mediated longevity. *Cell* **124**, 1039–1053 (2006).
47. Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J. & Vanfleteren, J. R. Selection and validation of a set of reliable reference genes for quantitative sod gene expression analysis in *C. elegans*. *BMC Mol. Biol.* **9**, 9 (2008).

HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle

Matthew A. Deardorff^{1,2*}, Masashige Bando^{3*}, Ryuichiro Nakato^{3*}, Erwan Watrin^{4*}, Takehiko Itoh⁵, Masashi Minamino³, Katsuya Saitoh³, Makiko Komata³, Yuki Katou³, Dinah Clark¹, Kathryn E. Cole⁶, Elfride De Baere⁷, Christophe Decroos⁶, Nataliya Di Donato⁸, Sarah Ernst¹, Lauren J. Francey¹, Yolanda Gyftodimou⁹, Kyotaro Hirashima¹⁰, Melanie Hullings¹, Yuuichi Ishikawa¹¹, Christian Jaulin⁴, Maninder Kaur¹, Tohru Kiyono¹², Patrick M. Lombardi⁶, Laura Magnaghi-Jaulin⁴, Geert R. Mortier¹³, Naohito Nozaki¹⁴, Michael B. Petersen^{9,15}, Hiroyuki Seimiya¹⁰, Victoria M. Siu¹⁶, Yutaka Suzuki¹⁷, Kentaro Takagaki¹⁸, Jonathan J. Wilde¹, Patrick J. Willems¹⁹, Claude Prigent⁴, Gabriele Gillesen-Kaesbach²⁰, David W. Christianson⁶, Frank J. Kaiser²⁰, Laird G. Jackson^{1,21}, Toru Hirota¹⁷, Ian D. Krantz^{1,2} & Katsuhiko Shirahige^{3,22}

Cornelia de Lange syndrome (CdLS) is a dominantly inherited congenital malformation disorder, caused by mutations in the cohesin-loading protein NIPBL^{1,2} for nearly 60% of individuals with classical CdLS^{3–5}, and by mutations in the core cohesin components SMC1A (~5%) and SMC3 (<1%) for a smaller fraction of probands^{6,7}. In humans, the multisubunit complex cohesin is made up of SMC1, SMC3, RAD21 and a STAG protein. These form a ring structure that is proposed to encircle sister chromatids to mediate sister chromatid cohesion⁸ and also has key roles in gene regulation⁹. SMC3 is acetylated during S-phase to establish cohesiveness of chromatin-loaded cohesin^{10–13}, and in yeast, the class I histone deacetylase Hos1 deacetylates SMC3 during anaphase^{14–16}. Here we identify HDAC8 as the vertebrate SMC3 deacetylase, as well as loss-of-function *HDAC8* mutations in six CdLS probands. Loss of HDAC8 activity results in increased SMC3 acetylation and inefficient dissolution of the ‘used’ cohesin complex released from chromatin in both prophase and anaphase. SMC3 with retained acetylation is loaded onto chromatin, and chromatin immunoprecipitation sequencing analysis demonstrates decreased occupancy of cohesin localization sites that results in a consistent pattern of altered transcription seen in CdLS cell lines with either *NIPBL* or *HDAC8* mutations.

Human SMC3 is acetylated by ESCO1 and ESCO2, homologues of yeast Eco1, and has been shown to be important for the establishment of sister chromatid cohesion^{10,11,13,17,18}. Using a monoclonal antibody specific for acetylated SMC3 (SMC3-ac)¹⁸, we found that whereas total SMC3 levels remain stable throughout the cell cycle, SMC3-ac rapidly disappears during mitosis, suggesting coordinated deacetylation (Supplementary Fig. 1).

We therefore used RNA interference (RNAi)-based screening of all known human histone deacetylases (HDACs) and sirtuins to identify HDAC8 as the vertebrate SMC3 deacetylase (Supplementary Fig. 2). Loss of HDAC8 activity using either *HDAC8* RNAi or the HDAC8-specific inhibitor PCI-34051 (PCI; Fig. 1a, b) does not alter cell cycle progression, but clearly increases SMC3-ac in both soluble and chromatin fractions throughout the cell cycle (Fig. 1c, lanes 4 and 6,

Fig. 1e, lanes 18–22 and 29–33, and Supplementary Fig. 3d, lanes 22–28 and 36–42). Nearly all HDAC8 is present in the soluble fraction in both asynchronous and synchronized cultures (Fig. 1c, e). These data indicate that HDAC8 is present and active throughout the cell cycle, and that soluble SMC3-ac is its deacetylation target, similar to Hos1 in yeast^{14–16}. Notably, the increase of SMC3-ac in the soluble fraction in the absence of HDAC8 activity suggests that SMC3-ac dissociates from chromatin but fails to be deacetylated. In addition, we unexpectedly observed few sister chromatid cohesion defects with loss of HDAC8 activity alone (Supplementary Fig. 4).

To understand the role of HDAC8 in genome-wide regulation of cohesin dynamics, we performed chromatin immunoprecipitation sequencing (ChIP-Seq) analysis of synchronized HeLa cells transfected with control or *HDAC8* RNAi (Fig. 2) and immunoprecipitated with either an anti-RAD21 antibody to detect total cohesin or an anti-SMC3-ac antibody. Although total cellular cohesin shows no decrease (Supplementary Fig. 5a, b), and there is a high degree of overlap between SMC3-ac, cohesin and CTCF¹⁹ localization sites in treated and untreated cells, high read numbers and tight correlations between experimental replicates enabled us to identify a 17% loss of total cohesin localization peaks with reduced HDAC8 activity (Fig. 2a–c and Supplementary Fig. 5d–h). Furthermore, despite using conditions that increase total SMC3-ac more than twofold (Fig. 1e and Supplementary Fig. 3d), we note a 16% loss of SMC3-ac localization sites with HDAC8 reduction (Fig. 2a–d and Supplementary Fig. 5f, g). Finally, we found that in both control and HDAC8-depleted cells, SMC3-ac preferentially localizes to downstream regions of genes relative to the distribution of RAD21 (Fig. 2c, d and Supplementary Fig. 5f, g). Together, this data demonstrates decreased occupancy of cohesin localization sites with the loss of HDAC8 activity, an observation similarly noted for *NIPBL* haploinsufficient CdLS cells²⁰.

With the known role of cohesin regulation in CdLS and the observations that a reduction in either HDAC8 or *NIPBL* leads to decreased cohesin occupancy of localization sites, we proposed that *HDAC8* mutations may cause CdLS. We screened this X-linked gene in 154 individuals with CdLS negative for mutations in *NIPBL*, *SMC1A* and

¹Division of Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, Pennsylvania 19104, USA. ²The Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania 19104, USA. ³Research Center for Epigenetic Disease, Institute for Molecular and Cellular Biosciences, The University of Tokyo, Tokyo 113-0032, Japan. ⁴Centre National de la Recherche Scientifique (CNRS), Research Institute of Genetics and Development (IGDR), Faculté de Médecine, Rennes 35043, France. ⁵School and Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8503, Japan. ⁶Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁷Center for Medical Genetics, Ghent University Hospital, 9000 Ghent, Belgium. ⁸Institut für Klinische Genetik, Technische Universität Dresden, 01307 Dresden, Germany. ⁹Department of Genetics, Institute of Child Health, 11527 Athens, Greece. ¹⁰Division of Molecular Biotechnology, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. ¹¹Department of Pathology, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. ¹²Virology Division, National Cancer Center Research Institute, Tokyo 104-0045, Japan. ¹³Department of Medical Genetics, Antwerp University Hospital and University of Antwerp, B-2650 Antwerp, Belgium. ¹⁴Bio-Frontier Research Center, Tokyo Institute of Technology, Yokohama 226-8503, Japan. ¹⁵Department of Clinical Genetics, Aalborg Hospital, Aarhus University Hospital, 9100 Aalborg, Denmark. ¹⁶Medical Genetics, University of Western Ontario, London, Ontario N6A 5W9, Canada. ¹⁷Graduate School of Frontier Sciences, The University of Tokyo, Tokyo 277-8561, Japan. ¹⁸Experimental Pathology, Japanese Foundation for Cancer Research, Tokyo 135-8550, Japan. ¹⁹GENDIA, 2020 Antwerp, Belgium. ²⁰Institut für Humangenetik Lübeck, Universität zu Lübeck, 23538 Lübeck, Germany. ²¹Department of Obstetrics and Gynecology, Drexel University School of Medicine, Philadelphia, Pennsylvania 19102, USA. ²²CREST, JST, K's Gobancho, 7, Gobancho, Chiyoda-ku, Tokyo 102-0076, Japan.

*These authors contributed equally to this work.

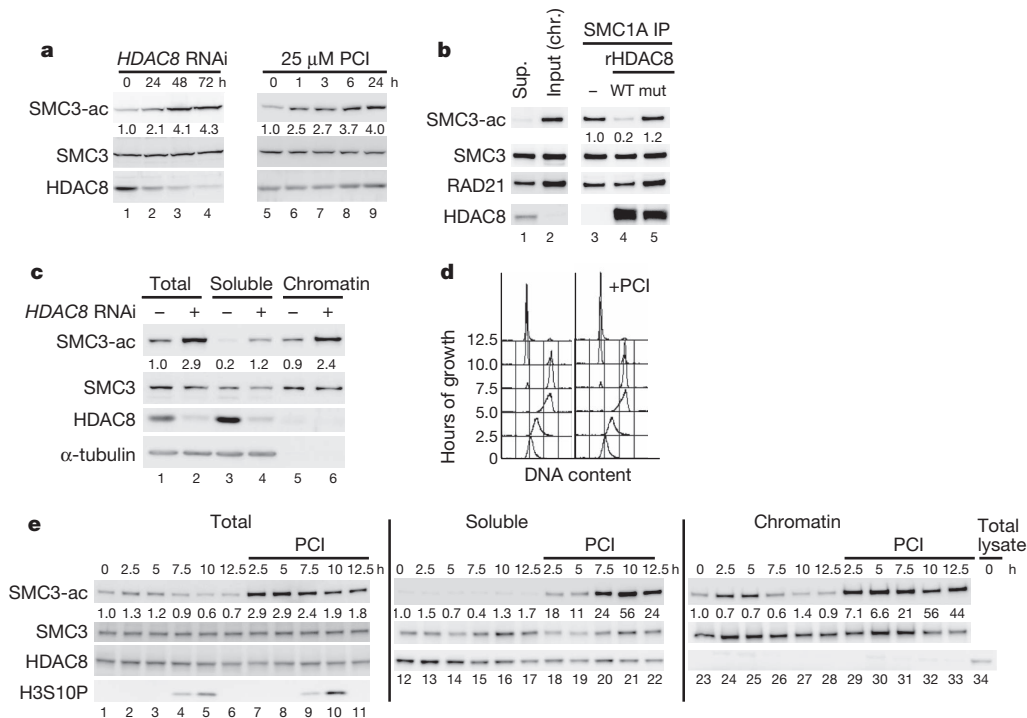


Figure 1 | HDAC8 is an SMC3 deacetylase. **a**, HeLa cells were transfected with *HDAC8* short interfering RNA (siRNA; lanes 1–4) or incubated with 25 μM PCI, the *HDAC8*-specific inhibitor (lanes 5–9), for the indicated times. Total cell lysates were prepared and analysed by immunoblotting using anti-SMC3-ac, SMC3 and *HDAC8* antibodies. Numbers beneath SMC3-ac bands indicate quantification of SMC3-ac levels normalized to SMC3 levels and the 0 h time point. **b**, Acetylated SMC3 was prepared by co-immunoprecipitation with SMC1A from HeLa cell chromatin extracts. Immunoprecipitates (IP) were incubated with recombinant purified *HDAC8* (rHDAC8) or mutant (mut) *HDAC8* protein at 30 °C for 1 h and analysed by immunoblotting as in **a**.

SMC3, as well as *RAD21* (also known as *SSC1*), *STAG2*, *ESCO1*, *ESCO2* and *MAU2* (also known as *SCC4*). We identified four *de novo* missense mutations and one *de novo* nonsense mutation in *HDAC8* (Supplementary Table 1 and Fig. 3a). In addition, one familial mutation (c.1001A>G; p.H334R) was identified in a boy, his mildly affected sister and his unaffected mother, in which the mutant allele was inactivated in her blood. This mutation was also one of the *de novo* mutations in an unrelated girl. None of the mutations was seen in 290 ethnically matched control chromosomes or in 629 individuals of the 1000 Genomes Project²¹. Despite the small numbers and the varied clinical features in females due to random X-chromosome inactivation, these children demonstrate growth, cognitive and facial features consistent with those caused by mutations in *NIPBL* ('classical' CdLS). Both expression studies (Supplementary Fig. 6a) and X-chromosome inactivation studies (data not shown) demonstrate complete skewing towards the normal allele in the blood of females with *HDAC8* mutations, indicating strong selection against the mutation. This limited the number of cell lines available for subsequent studies. However, immunoblotting demonstrated minimal *HDAC8* protein expression in lymphoblastoid cell lines (LCLs) from a hemizygous boy with a p.G320R mutation as well as the skin fibroblasts from a female lyonized to express a p.H180R mutant allele (Fig. 3b), indicating protein instability in each case. Consistent with this, assessment of SMC3-ac demonstrates increased levels in both the p.G320R LCLs and the p.H180R fibroblasts, whereas the total amount of SMC3 (Fig. 3b) and cell cycle distribution for the LCLs are unchanged (Supplementary Fig. 6b).

We expressed the *HDAC8* missense mutations in *Escherichia coli*, purified and assayed each for deacetylase activity. These data (Fig. 3c) demonstrate that the p.H180R mutation severely abrogates *HDAC8* deacetylase activity, and demonstrate significant activity losses for

the p.G320R, p.T311M and p.H334R mutations, consistent with high conservation of each residue and position in the structure of *HDAC8*-substrate complexes^{22,23} (Fig. 3d and Supplementary Fig. 6c–e). Finally, purified wild-type, but not mutant *HDAC8*, can rescue the SMC3 overacetylation seen in *HDAC8* mutant LCLs (Supplementary Fig. 6g, h).

We have previously demonstrated that mutations in *NIPBL*, the primary cause of CdLS, result in consistent, reproducible changes of genes expressed from LCLs derived from individuals with CdLS²⁰. Using Nanostring multiplex expression analysis to avoid variation introduced by PCR or RNA amplification-based assessments and a previously validated 32-gene CdLS classifier set²⁰, we compared the expression of LCLs from the male with the p.G320R mutation and from two girls (with the p.H180R and p.H334R mutations) who express only the normal allele in these cells, to 10 normal controls and 12 LCLs with loss of function *NIPBL* mutations. In this assay, the aggregate 32-gene LCL expression profile of the single male with an *HDAC8* mutation strongly correlates with that seen in *NIPBL*-mutant cell lines, whereas the profile for the two female lines, which express the wild-type allele, correlates with normal controls (Fig. 3e). These data support the hypothesis that loss of *HDAC8* activity results in widespread transcriptional dysregulation as seen in *NIPBL*-mutated CdLS cells. The loss of cohesin-binding sites in *HDAC8* RNAi-depleted HeLa cells and in *NIPBL*-mutant LCLs²⁰ was also noted in the p.G320R *HDAC8*-mutant LCLs and fibroblasts derived from the girl with the p.H180R mutation, in which small but consistent changes in transcriptional dysregulation were also noted (Supplementary Figs 6a, 7 and 8). Together, these data suggest that loss of *HDAC8* activity leads to a common pathogenic mechanism for classical CdLS that converges on the reduction of bound cohesin complexes.

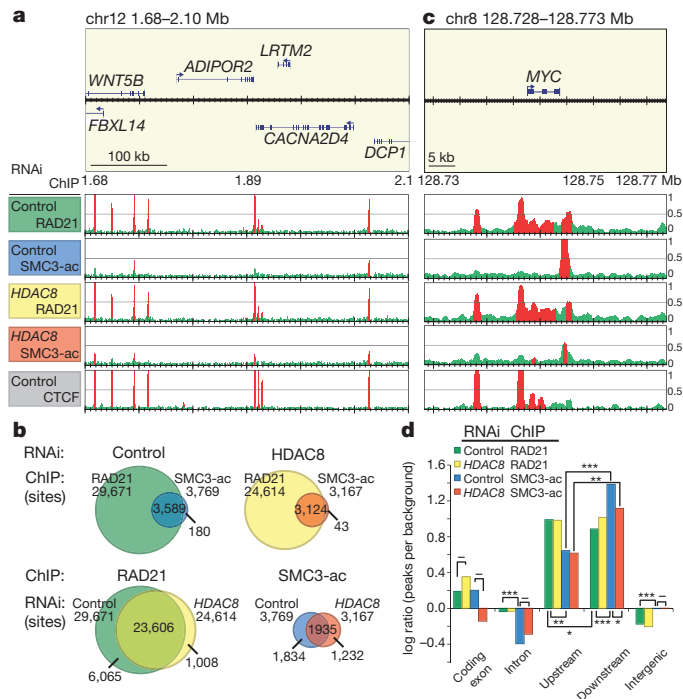


Figure 2 | Cohesin and SMC3-ac localization sites in control and HDAC8 RNAi-treated HeLa cells. **a**, Example of ChIP-Seq data (Ensemble gene position 1.68–2.10 Mb of human chromosome 12). ChIP-Seq data are shown in reads per million. Regions in which signals were significantly enriched (see Methods) are in red. Binding profiles for RAD21 and SMC3-ac in control and HDAC8 RNAi-treated HeLa cells are shown. **b**, Venn diagrams showing numbers and overlap of localization sites. **c**, Example of region-specific localization showing binding sites for RAD21 and SMC3-ac in the MYC gene locus. Ensemble gene position 128.728–128.773 Mb of human chromosome 8 is shown. **d**, Classification of RAD21 and SMC3-ac localization sites as a log ratio to background reads. Significance was calculated using a two-sample test for equality of proportions. * $P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$; – $P > 0.05$. Upstream and downstream are defined as within 5 kb from gene boundaries, respectively. SMC3-ac preferentially localizes to the downstream end of genes.

To understand the endogenous effect of HDAC8 loss on cohesin further, we compared each cohesin subunit in chromatin-bound and unbound fractions in the LCLs with a hemizygous p.G320R mutation versus control LCLs, and unexpectedly noted accumulation of RAD21 fragments in HDAC8-mutant cells (Supplementary Fig. 9a, lanes 2, 4 and 6). Using a RAD21 amino terminal-specific antibody, we verified accumulation of the RAD21 N-terminal separate cleavage fragment (RAD21-N) in mutant LCLs, HDAC8 RNAi-treated HeLa cells and HDAC8-mutated fibroblasts expressing catalytically inactive HDAC8 compared with wild-type HDAC8 (Fig. 4 and Supplementary Fig. 9). The accumulation of this cleavage fragment in the chromatin-bound fraction of HDAC8-mutant cells suggests that it may remain bound to the cohesin complex, so we assessed whether RAD21-N co-immunoprecipitates with other cohesin subunits. In HDAC8-mutant cells, RAD21-N co-immunoprecipitates with SMC1A, SMC3, STAG1 and STAG2 (Fig. 4b), strongly suggesting that without functional HDAC8, the cleaved N terminus of RAD21 remains attached to the cohesin complex, presumably via acetylated SMC3 (ref. 24). To confirm that clearance of cleaved RAD21 is dependent on HDAC8, we performed complementary analyses in synchronized HeLa cells with or without the HDAC8 inhibitor PCI (Fig. 4c, d). These data demonstrate the appearance of the N-terminal and carboxy-terminal RAD21 fragments after entry into mitosis (marked by histone H3 Ser10 phosphorylation; H3S10P) and activation of separase (Fig. 4d, lanes 9–12 and 21–24) as well as their persistence in both soluble and chromatin fractions (Fig. 4d, lanes 18–20) when most cells are in G1

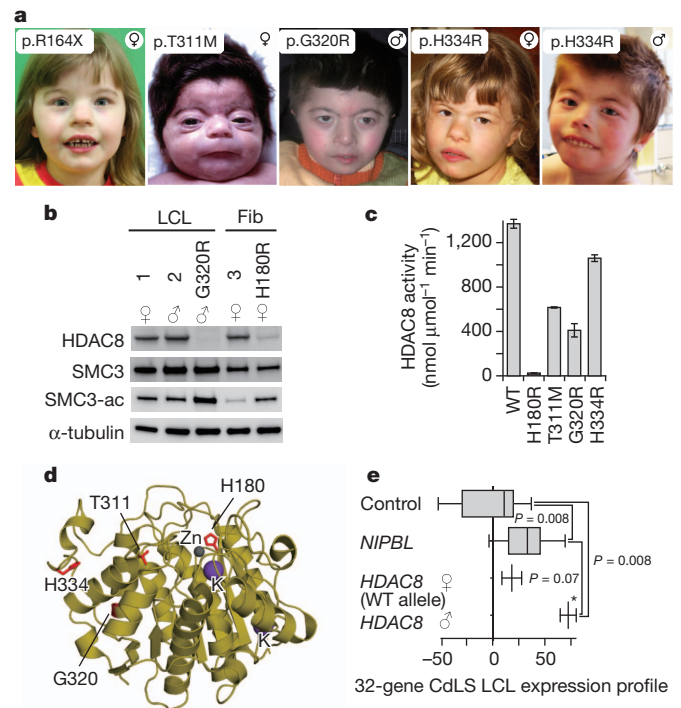


Figure 3 | HDAC8 mutations in CdLS. **a**, Facial features of individuals with HDAC8 mutations labelled with corresponding mutation and sex. **b**, Western blotting of protein from LCLs and fibroblasts (Fib). Sex is indicated under controls (1, 2 and 3) or mutation designation. **c**, HDAC8 mutations disrupt deacetylase activity. Bar graphs demonstrate the effect of HDAC8 mutations on deacetylase specific activity (nmol substrate $\mu\text{mol enzyme}^{-1} \text{min}^{-1}$). All assays were performed in triplicate. Error bars indicate standard deviation. Enzymatic activity for all mutations is significantly less than wild type ($P < 0.05$). **d**, Localization of HDAC8 mutations on the crystal structure (PDB accession code 3F06; ref. 23). Mutated residues are in red. **e**, The expression profile of HDAC8-mutant LCL is consistent with that seen in cell lines with NIPBL mutations. The summed standard deviations of gene expression compared with NIPBL mutant from control LCLs for 32 genes was determined for each sample. Box graphs demonstrate twenty-fifth, median and seventy-fifth percentiles of each group. Whiskers indicate minimum and maximum values. Note that for the HDAC8 male (p.G320R) the values are for biological replicate samples for a single cell line (denoted by asterisk). Unpaired two-tailed t -test significances are indicated.

phase (data not shown). By contrast, in untreated cells, these cleavage products are predominantly detected in the soluble fraction (Fig. 4d, lanes 2–4), and largely disappear 3 h after release.

We also made several observations about HDAC8 activity in the prophase removal of cohesin^{25,26}. Loss of HDAC8 activity does not disrupt the prophase pathway of cohesin removal from chromatin, as demonstrated by a consistent reduction of chromatin-bound cohesin in the presence of metaphase inhibitors in either control or PCI-treated cells (Fig. 4d, lanes 13 and 17 versus 21–24). However, HDAC8 activity is necessary after prophase removal of cohesin, because in metaphase inhibitor-treated cells, soluble SMC3-ac is markedly increased in the presence of PCI (Fig. 4d, lanes 9–12). Finally, with WAPAL depletion, which diminishes prophase removal of cohesin^{18,27,28}, we noted that SMC3-ac is substantially increased (Supplementary Fig. 9e).

To test whether other cohesin-promoting proteins might remain associated with SMC3-ac in the absence of HDAC8 activity, we analysed HeLa and patient-derived cells, which also demonstrated increased retention of sororin (also known as CDCA5)^{18,29} and RAD21-N, but not WAPAL, on soluble SMC3-ac (Fig. 4e–g and Supplementary Fig. 9d). Taken together, these data strongly suggest that HDAC8 is necessary for the deacetylation of SMC3 after its removal from chromatin in both prophase and anaphase pathways to enable proper

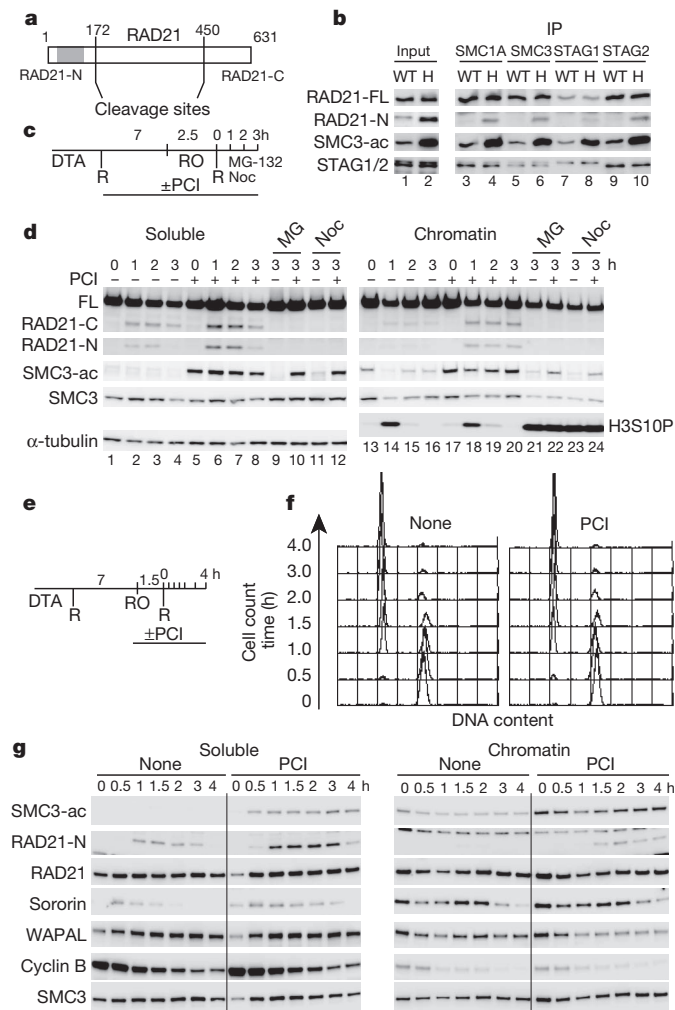


Figure 4 | Retention of RAD21-N and sororin on cohesin in the absence of HDAC8. **a**, RAD21 separase cleavage sites. Vertical lines indicate cleavage sites. Grey region indicates the RAD21-N antibody epitope. **b**, SMC3, STAG1 and STAG2 were immunoprecipitated from soluble extracts from normal (WT) and HDAC8-mutant (H) cell lines. Full-length RAD21 (FL), the cleaved N-terminal RAD21 fragment, SMC3-ac and STAG1/2 were analysed by SDS-PAGE and immunoblotting. **c**, Experimental schema of **d**. HeLa cells were synchronized by double thymidine arrest (DTA) and released (R) in the presence or absence of PCI. Seven hours after release, cells were treated with RO-3306 (RO) to arrest at the G2/M boundary. After removing RO-3306, cells were cultured in the presence or absence of nocodazole (Noc) or MG-132 (MG) to arrest in metaphase. Cells were collected at indicated time points after release from RO-3306 arrest. **d**, Cell extracts were fractionated into soluble and chromatin-bound fractions. Full-length RAD21, cleaved C- and N-terminal RAD21 fragments, SMC3-ac, SMC3, H3S10P and α -tubulin were analysed by SDS-PAGE and immunoblotting. H3S10P is used as a metaphase marker. **e**, Experimental schema of **f** and **g**. HeLa cells were synchronized and released and treated with RO-3306 as in **d**. After removing RO-3306, cells were cultured in the presence or absence of PCI and collected at the indicated times. **f**, FACS analysis of HeLa cells in **g**. Cell extracts were fractionated into soluble and chromatin-bound fractions. Levels of SMC3-ac, RAD21-N, full-length RAD21, sororin, WAPAL, cyclin B (a G2 marker) and SMC3 were analysed by SDS-PAGE and immunoblotting.

dissolution of pro-cohesive elements and allow for recycling of 'refreshed' cohesin for the next cell cycle (Supplementary Fig. 10).

In summary, this work demonstrates that HDAC8 functions as a vertebrate SMC3 deacetylase to facilitate the renewal of cohesin after its removal from chromatin in prophase or anaphase, and that loss of HDAC8 activity results in decreased cohesin at localized sites to cause both cellular and clinical features of CdLS.

METHODS SUMMARY

Cell culture and synchronization, chromatin fractionation, purification of recombinant HDAC8 in baculovirus and *E. coli*, SMC3-ac deacetylation assays, RNA interference, immunofluorescence microscopy, human subjects, mutation screening, HDAC8 conservation analysis, crystal structure analysis, expression profiling, assessment of X-chromosome inactivation, chromatin immunoprecipitation, ChIP-Seq and RNA-Seq analyses are described in detail in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 20 January; accepted 12 June 2012.

Published online 12 August 2012.

- Gillespie, P. J. & Hirano, T. Scc2 couples replication licensing to sister chromatid cohesion in *Xenopus* egg extracts. *Curr. Biol.* **14**, 1598–1603 (2004).
- Takahashi, T. S., Yiu, P., Chou, M. F., Gygi, S. & Walter, J. C. Recruitment of *Xenopus* Scc2 and cohesin to chromatin requires the pre-replication complex. *Nature Cell Biol.* **6**, 991–996 (2004).
- Gillis, L. A. *et al.* NIPBL mutational analysis in 120 individuals with Cornelia de Lange syndrome and evaluation of genotype-phenotype correlations. *Am. J. Hum. Genet.* **75**, 610–623 (2004).
- Krantz, I. D. *et al.* Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nature Genet.* **36**, 631–635 (2004).
- Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J. & Strachan, T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nature Genet.* **36**, 636–641 (2004).
- Deardorff, M. A. *et al.* Mutations in cohesin complex members SMC3 and SMC1A cause a mild variant of Cornelia de Lange syndrome with predominant mental retardation. *Am. J. Hum. Genet.* **80**, 485–494 (2007).
- Musio, A. *et al.* X-linked Cornelia de Lange syndrome owing to SMC1L1 mutations. *Nature Genet.* **38**, 528–530 (2006).
- Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
- Dorsett, D. Cohesin: genomic insights into controlling gene transcription and development. *Curr. Opin. Genet. Dev.* **21**, 199–206 (2011).
- Rolef Ben-Shahar, T. *et al.* Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science* **321**, 563–566 (2008).
- Unal, E. *et al.* A molecular determinant for the establishment of sister chromatid cohesion. *Science* **321**, 566–569 (2008).
- Zhang, B. *et al.* Dosage effects of cohesin regulatory factor PDS5 on mammalian development: implications for cohesinopathies. *PLoS ONE* **4**, e5232 (2009).
- Heidinger-Pauli, J. M., Unal, E. & Koshland, D. Distinct targets of the Eco1 acetyltransferase modulate cohesin in S phase and in response to DNA damage. *Mol. Cell* **34**, 311–321 (2009).
- Beckouët, F. *et al.* An SMC3 acetylation cycle is essential for establishment of sister chromatid cohesion. *Mol. Cell* **39**, 689–699 (2010).
- Xiong, B., Lu, S. & Gerton, J. L. Hos1 is a lysine deacetylase for the Smc3 subunit of cohesin. *Curr. Biol.* **20**, 1660–1665 (2010).
- Borges, V. *et al.* Hos1 deacetylates Smc3 to close the cohesin acetylation cycle. *Mol. Cell* **39**, 677–688 (2010).
- Zhang, J. *et al.* Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. *Mol. Cell* **31**, 143–151 (2008).
- Nishiyama, T. *et al.* Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* **143**, 737–749 (2010).
- Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
- Liu, J. *et al.* Transcriptional dysregulation in NIPBL and cohesin mutant human cells. *PLoS Biol.* **7**, e1000119 (2009).
- The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
- Vannini, A. *et al.* Substrate binding to histone deacetylases as shown by the crystal structure of the HDAC8-substrate complex. *EMBO Rep.* **8**, 879–884 (2007).
- Dowling, D. P., Gantt, S. L., Gattis, S. G., Fierke, C. A. & Christianson, D. W. Structural studies of human histone deacetylase 8 and its site-specific variants complexed with substrate and inhibitors. *Biochemistry* **47**, 13554–13563 (2008).
- Haering, C. H., Lowe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, 773–788 (2002).
- Waizenegger, I. C., Hauf, S., Meinke, A. & Peters, J. M. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* **103**, 399–410 (2000).
- Giménez-Abian, J. F. *et al.* Regulation of sister chromatid cohesion between chromosome arms. *Curr. Biol.* **14**, 1187–1193 (2004).
- Gandhi, R., Gillespie, P. J. & Hirano, T. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr. Biol.* **16**, 2406–2417 (2006).
- Kueng, S. *et al.* Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955–967 (2006).
- Schmitz, J., Watrin, E., Lenart, P., Mechtler, K. & Peters, J. M. Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Curr. Biol.* **17**, 630–636 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to the individuals and families with Cornelia de Lange syndrome who participated in this study, as well as to the referring physicians and colleagues who have contributed samples and clinical information. We thank Riken Omics Science Center, K. Nakagawa, S. Watanabe, M. Albrecht and J. Eckhold for technical support. We thank J.-M. Peters for the sororin and RAD21 antibodies. We thank F. Beckouët and K. Nasmyth for sharing unpublished results. We are indebted to the continued support of the US and International Cornelia de Lange Syndrome Foundations. This work was supported by National Institutes of Health grants K08HD055488 (NICHD, M.A.D.), GM49758 (D.W.C.) and P01 HD052860 (NICHD; I.D.K.), research grants from the US CdLS Foundation, institutional funds from the Children's Hospital of Philadelphia, Intramural funding from the University of Lübeck (Schwerpunktprogramm, Medizinische Genetik: Von seltenen Varianten zur Krankheitsentstehung; F.J.K., G.G.-K.), Research Program of Innovative Cell Biology by Innovative Technology, Grant-in-Aid for Scientific Research (S) and for innovative science from MEXT (K.Sh.).

Author Contributions M.B. and K.Sh. designed and performed the biochemical analyses; M.B., E.W., C.J., L.M.-J. and K.Sh. performed HDAC screening. M.B., K.Sa., E.W. and L.M.-J. performed chromatin spreading assays; N.N. monoclonal

isolation; M.B., R.N., T.I., M.K., Y.K. and K.Sh. the ChIP sequencing and analysis; M.B., R.N. and Y.S. RNA sequencing; M.M., K.T. and T.H. microscopy. M.A.D., G.G.-K., L.G.J., F.J.K. and I.D.K. initiated the human studies. M.A.D., D.C., E.D.B., G.R.M., V.M.S., P.W., N.T., Y.G., M.B.P., G.G.-K. and I.D.K. identified and characterized clinical data. M.A.D., M.K., S.E., L.J.F., M.H. and F.J.K. performed mutation screening, inactivation and lymphoblastoid expression studies. K.H., T.K. and H.S. established fibroblast cell lines. M.A.D., J.J.W., K.E.C., P.M.L., C.D. and D.W.C. performed enzymatic and structural analysis. M.A.D., M.B. and K.Sh. drafted the manuscript. All authors analysed data, discussed the results and commented on the manuscript.

Author Information ChIP-Seq and RNA-Seq data from this study are available from the Sequence Read Archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>) under the accession number SRP011927. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A.D. (deardorff@email.chop.edu), I.D.K. (ian2@mail.med.upenn.edu) or K.Sh. (kshirahi@iam.u-tokyo.ac.jp).

METHODS

Antibodies and reagents. Primary antibodies used in this study are as previously described for RAD21 (hSCC1)¹⁹, SMC3-ac¹⁸ and the RAD21 N terminus³⁰. Antibodies used are as follows: SA1, SA2, SMC1A, HDAC6, HDAC10, SIRT4, SIRT5, SIRT7 were from Abcam; α -tubulin was from Sigma; histone H3(S10P), HDAC1, HDAC2, HDAC3, HDAC4, HDAC5, SIRT1, SIRT3, SIRT6, SIRT7 and MAD2L1 were from Cell Signaling Technology; HDAC7, HDAC11 and SIRT2 were from Millipore; and SMC3 was from Bethyl Laboratories. HDAC8 antibodies were purchased from Santa Cruz Biotechnology or derived as mouse monoclonal HDAC8 antibody raised against peptide CDAYLQHLQKVSQEGDDHPDS. PCI-34051 was purchased from Sigma Chemical or synthesized by Synstar Japan. TSA, sodium butyrate and nicotinamide were from Wako Chemical. Suberoyl anilide bishydroxamide was from Toronto Research Chemicals. Nocodazole was from Calbiochem.

Cell culture and synchronization. HeLa cells were cultured in DMEM (Invitrogen) supplemented with 0.2 mM L-glutamine, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and 10% FCS. LCLs were cultured in RPMI-1640 supplemented with 0.2 mM L-glutamine, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and 20% FCS. Fibroblasts were cultured in DMEM (Invitrogen) supplemented with 0.2 mM L-glutamine, 100 U ml⁻¹ penicillin, 100 μ g ml⁻¹ streptomycin and 20% FCS. For cell cycle synchronization, HeLa cells were synchronized by double thymidine arrest (14–16 h in the presence of 2 mM thymidine, 8 h release, 16 h in the presence of 2 mM thymidine), and collected 6 h and 13 h after the second release for enrichment in the G2 and G1 phase, respectively. For enrichment in prometaphase of mitosis, nocodazole (330 nM) was added 8 h after release from the second thymidine block for 1 or 2 h, and cells were collected by shake off.

Protein extraction and immunoprecipitations. To obtain total cell extracts, cells were lysed with lysis buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM KCl, 5 mM MgCl₂, 0.2% NP-40, 10% glycerol, cOmplete protease inhibitor cocktail (Roche), and phosSTOP phosphatase inhibitor (Roche)) and DNA in the chromatin fraction was digested by the treatment of benzonase (Novagen). To obtain soluble cell extracts, the cells were lysed with lysis buffer containing 200 mM NaCl and centrifuged at 20,000g. Immunoprecipitations were performed using ProteinG magnetic beads conjugated with 5 μ g antibodies and incubated with the cell extracts for 5 h at 4 °C. After washing with lysis buffer, the beads were suspended with SDS-PAGE sample buffer.

RNA isolation and RT-PCR. Total RNA was isolated using Trizol (Invitrogen) and Nucleospin RNA II (Macherey-Nagel) following the manufacturer's instructions. Synthesis of complementary DNA was performed using SuperScript III first-strand synthesis system (Invitrogen). cDNA was amplified with KAPA SYBR fast qPCR kit (Kapa Biosystems) on an ABI 7500 real time PCR system (Applied Biosystems). Primer sequences were: HDAC9 forward, 5'-CTTTGTC AGGTTCTGCTGTTCTC-3', reverse, 5'-TTCCTGTTCCACAAGGCATTTC-3'; ACTB forward, 5'-TGGCACCAGCACAATGAA-3', reverse, 5'-CTAAGTCA TAGTCCGCTAGAAGCA-3'; and GAPDH forward, 5'-GCACCGTCAAGG CTGAGAAC-3', reverse, 5'-TGGTGAAGACGCCAGTGA-3'.

Immunofluorescence microscopy. HeLa cells grown on 18-mm coverslips were fixed with 4% paraformaldehyde and permeabilized with PBS containing 0.5% Triton X-100 for 10 min. After blocking with goat serum in PBS, cells were incubated with primary antibodies for SMC3 or SMC3-ac. DNA was stained with 4',6-diamidino-2-phenylindole (DAPI).

Chromatin fractionation. Performed as previously described³¹.

Chromosome spreads. Cells treated with nocodazole (100 ng ml⁻¹) were collected by mitotic shake off and hypotonically swollen in 40% PBS, 60% tap water for 5 min at room temperature. Cells were fixed with Carnoy's solution (methanol:acetic acid = 3:1), dropped on glass slides and dried. Slides were stained with 5% Giemsa (Merck), and washed with water, air-dried and mounted with Entellan (Merck).

Purification of recombinant HDAC8 and SMC3-ac deacetylation assay. Full-length cDNA of HDAC8 and the enzymatically dead HDAC8 mutant (D101A/Y306F)²² was subcloned into pFASTBacHT (Invitrogen). The Bac-to-Bac baculovirus expression system (Invitrogen) was used to produce recombinant virus according to the manufacturer's protocol. To express recombinant protein, Sf9 cells were infected with each baculovirus and incubated for 60 h at 27 °C. Total cell lysates were prepared using Laemmli SDS buffer and boiling for 5 min. For purification of HDAC8 protein, infected cells were lysed with buffer A (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.1% NP-40 and 10 mM 2-mercaptoethanol (2-ME)). After centrifugation at 3,000g the supernatants were incubated with Ni-NTA beads for 2 h at 4 °C. The beads were washed with buffer A, washed with buffer B (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10% glycerol 0.05% NP-40 and 1 mM dithiothreitol) and eluted with buffer B containing various concentration of imidazole (50, 100, 200 or 350 mM). The elution fractions including HDAC8

protein were collected and concentrated using a Centricon concentrator. For detection of SMC3 deacetylation by HDAC8, solubilized total lysates treated with benzonase or SMC1A immunoprecipitants were used as substrates. Substrates were incubated with recombinant HDAC8 or HDAC8 mutant in reaction buffer (50 mM HEPES-potassium hydroxide, pH 7.4, 10% glycerol, 10 mM KCl, 2.5 mM MgCl₂, 100 mM NaCl, 0.1 mM phenylmethylsulfonyl fluoride, 50 ng ml⁻¹ BSA and 0.02% NP-40) for 1 h at 30 °C and then boiled with SDS sample buffer.

RNA interference. The siRNA oligonucleotides were annealed according to manufacturer's instruction and used at a final concentration of 100–200 nM. The siRNA transfections were performed using oligofectamine (Invitrogen). Oligonucleotides targeting the firefly luciferase GL2 were used as controls. All siRNA sequences are listed in Supplementary Table 2.

Human subjects. All individuals enrolled in the study were diagnosed by clinical geneticists experienced in the diagnosis of CdLS to have clinical features consistent with a diagnosis of CdLS. All patients and family members were enrolled in the study under an Institutional Review Board-approved protocol of informed consent at The Children's Hospital of Philadelphia or the Institut für Humangenetik Lübeck.

Mutation screening. All genes discussed were screened for mutations in the coding exons and intron–exon boundaries using PCR of genomic DNA followed by sequencing. Primers were designed using ExonPrimer (<http://ihg.gsf.de/ihg/ExonPrimer.html>). Primer sequences and PCR conditions are available on request. Sequencing was performed using BigDye Terminator v3.1 cycle sequencing and analysed on an ABI 3730 (Applied Biosystems). All probands were pre-screened and were negative for mutations in NIPBL, SMC1A and SMC3.

Reference sequences and HDAC8 conservation analysis. National Center for Biotechnology Information (NCBI) RefSeq accessions for messenger RNAs and proteins, respectively, referenced in this work include: HDAC8 (NM_018486 and NP_060956), NIPBL (NM_133433 and NP_597677) and SMC1A (NM_006306 and NP_006297). HDAC8 protein sequences for *Homo sapiens* (AAF73428), *Bos taurus* (DAA12953), *Rattus norvegicus* (AAI62023), *Mus musculus* (CAM17598), *Danio rerio* (NP_998596), *Xenopus laevis* (NP_001085711), *Xenopus tropicalis* deduced from mRNA BC161282, *Drosophila melanogaster* (AAC61494) and for the *Saccharomyces cerevisiae* HDAC8-like proteins Hos2 (NP_011321) and Rpd3 (AAT92832) were aligned by the ClustalW method³² using MacVector software (Accelrys Corp).

Mapping mutations to the HDAC8 crystal structures. The identified HDAC8 mutations were mapped onto the crystal structure data of human HDAC8–substrate complex (PDB accession 3F06 (ref. 23)) to visualize location with respect to active domains of the protein using Cn3D³³, and PyMol³⁴ software.

Expression analysis of CdLS 32-gene classifier. Lymphocyte culture and RNA extraction were performed as previously described²⁰. Genes and Nanostring³⁵ probe sequences are listed in Supplementary Table 3. For each sample, 100 ng of RNA was analysed. Hybridizations were carried out by mixing 5 μ l of each RNA sample (normalized to 20 ng μ l⁻¹) with 20 μ l of NanoString nCounter Reporter probes and 5 μ l of nCounter Capture probes (30 μ l total reaction volume) and incubating the hybridizations at 65 °C for 18 h. After hybridization purification, NanoString reporter capture, stretching and imaging was performed as described³⁵. To account for differences in assay efficiency the data was normalized to the sum of six positive internal control RNA spikes that ranged from 0.125 to 128 fM. Internally normalized data were then subsequently normalized to the geometric mean of the ACTB and RPL19 spike-normalized counts. To assess reproducibility, Pearson's *r* correlation coefficient was calculated for three technical replicates (average *r* = 0.99986) and 11 biological replicates (average *r* = 0.99675). To define a 'CdLS' profile, average expression and standard deviations were calculated for 32 genes using RNA from LCLs from 10 normal controls and 12 NIPBL truncating or nonsense mutations. For each patient gene expression assay, the number of deviations in expression from normal was determined and divided by the average NIPBL expression standard deviation. These products were then summed for each gene to give a score representing a sum of deviation from normal in a CdLS manner for all 32 genes tested for each patient. The data and calculation results for each step are in Supplementary Data File 1.

Recombinant HDAC8 protein expression and purification from E. coli. Identified human HDAC8 mutations were introduced into a previously described HDAC8-6His-pET20b construct³⁶ using a QuickChange site-directed mutagenesis kit (Agilent Genomics). Oligonucleotide sequences are available on request. HDAC8 was recombinantly expressed in BL21(DE3) *E. coli* cells and purified over Talon resin (Clontech) according to published procedures²³, with minor modifications. In brief, 50 ml cultures (Luria–Bertani media supplemented with ampicillin (50 μ g ml⁻¹)) were grown overnight and used to inoculate 1-l flasks (minimal media supplemented with ampicillin at 50 μ g ml⁻¹). Cells were grown at 37 °C until *A*_{600 nm} of ~0.5, at which point the cells were induced by the addition of isopropyl β -D-thiogalactopyranoside (0.4 mM final concentration) and ZnCl₂

(100 μ M final concentration), and grown overnight at 18 °C. The cells were pelleted by centrifugation, resuspended in 25 ml of lysis buffer (50 mM Tris, pH 8.0, 500 mM KCl, 5% glycerol, 3 mM 2-ME and 115 μ M phenylmethanesulphonyl fluoride), and lysed by sonication on ice. Cell debris was pelleted by centrifugation and the cell-free extract was purified by affinity chromatography (Talon resin, Clontech Labs) using a step gradient (buffer A: 50 mM Tris, pH 8.0, 500 mM KCl, 5% glycerol and 3 mM 2-ME; buffer B: 50 mM Tris, pH 8.0, 500 mM KCl, 5% glycerol, 3 mM 2-ME and 250 mM imidazole). The final yield was approximately 8.6 mg l⁻¹ culture for wild-type HDAC8, 6.7 mg l⁻¹ for HDAC8(G320R), 6.1 mg l⁻¹ for HDAC8(H334R), 5.8 mg l⁻¹ for HDAC8(H180R) and 3.9 mg l⁻¹ for HDAC8(T311M); protein concentrations were determined using the colorimetric Bradford assay³⁷. The final purity of each protein sample was better than 90% as determined by SDS-PAGE analysis.

HDAC8 activity was measured using the commercially available Fluor-de-Lys HDAC8 deacetylase substrate and Developer II (Enzo Life Sciences). All assays were run at 25 °C and contained 0.5 μ M enzyme and 150 μ M substrate in assay buffer (25 mM Tris, pH 8.2, 137 mM NaCl, 2.7 mM KCl and 1 mM MgCl₂). After 30 min, the reactions were quenched by addition of the HDAC8 inhibitor M344 (Sigma Aldrich, 100 μ M final concentration) and Developer II. Fluorescence was measured using a Fluoroskan II plate reader (excitation = 355 nm, emission = 460 nm). Product concentration was calculated from raw data using standard curves for substrate and product (aminomethylcoumarin, Enzo Life Sciences), with the total concentration of substrate and product kept at 150 μ M (before dilution with the developer solution). All activity assays were performed in triplicate. Enzyme activities are reported as nmol product μ mol enzyme⁻¹ min⁻¹.

Assessment of X-chromosome inactivation. X-chromosome inactivation (XCI) was determined by evaluating the methylation status of the CAG microsatellite locus at the 5' end of the androgen receptor gene as previously described with modifications^{38,39}. In brief, genomic DNA was isolated from peripheral blood and two reaction digests were performed for each patient. In one reaction, 1 μ g of DNA was digested in 25 μ l with the methylation-sensitive restriction enzyme HpaII (New England Biolabs), which cuts the active unmethylated allele. In the other reaction, DNA was incubated in enzyme digest buffer without enzyme. After a 16 h incubation at 37 °C, digestion was terminated by incubation at 65 °C for 20 min. From each reaction, 2 μ l was then amplified by PCR with primers flanking the polymorphic androgen receptor CAG repeat as described³⁸; forward, 5'-CTGTGAAGGTTGCTGTTCCTCAT-3', reverse, 5'-FAM-TCCAGAATCTGTTCCAGAGCGTGC-3'. ABI3730 Genetic Analyzer and GeneMapper V4.0 software (Applied Biosystems) were used for genotyping analysis. The percentage of X-chromosome inactivation was calculated by dividing the ratio of the allele peak volumes in the HpaII-treated sample by the ratio of the allele peak volumes in the untreated sample. We considered alleles separated by more than two CAG repeats as informative⁴⁰ and used a cutoff of >80:20% for skewed XCI and >95:5% for extremely skewed XCI.

Fibroblast immortalization and lentiviral transduction. CDL016 fibroblasts were immortalized by exogenous expression of hTERT. HDAC8 and the enzymatically dead HDAC8 mutant (D101A/Y306F)²² were cloned into pENTR H1 Gateway vector and recombined with lentivirus vector, pCS-RfA-CG, using LR Clonase II (Invitrogen). Recombinant lentiviruses were produced from 293FT cells transfected with CS-CMV-HDAC8 or CS-CMV-HDAC8mut and packaging vectors. The fibroblasts were incubated with recombinant lentiviruses in the presence of 8 μ g ml⁻¹ polybrene and transduction efficiency was measured using green fluorescent protein (GFP) expression.

ChIP and quantitative PCR. Cells were crosslinked with 1% formaldehyde for 10 min, quenched with 125 mM glycine, and prepared for ChIP as previously described¹⁹. ChIP was performed as previously described using RAD21, SMC3-ac and control antibodies¹⁹. In brief, crosslinked cell lysates solubilized by sonication were incubated with protein A or protein G Dynabeads (Dyna), crosslinked with the antibodies for 14 h at 4 °C. After this, beads were washed several times and eluted with elution buffer (50 mM Tris, 10 mM EDTA and 1% SDS) for 20 min at 65 °C. The eluates were incubated at 65 °C overnight to reverse crosslinks and then treated with RNaseA and then with proteinase K. The samples were further purified by phenol-chloroform extraction and an extra purification step using a PCR purification kit (Qiagen). Quantitative PCR was performed using KAPA SYBR Fast qPCR kit (Kapa Biosystems) on an ABI 7500 PCR system (Applied Biosystems). The results were presented as a percentage of input chromatin that was precipitated. Primers used in this study were listed in Supplementary Table 4.

ChIP-Seq analyses. DNA from whole cell extracts (WCE) and ChIP fractions was further sheared to an average size of approximately 150 base pairs (bp) by ultrasonication (Covaris), end-repaired, ligated to sequencing adapters and

amplified according to manufacturer's instructions (Applied Biosystems SOLiD Library Preparation Protocol). Gel-purified amplified DNA between 100 and 150 bp was sequenced on the Applied Biosystems SOLiD platforms (SOLiD 3 and 5500) to generate single-end 50-bp reads. Sequenced reads of both ChIP fractions and WCE were aligned to the human genome (UCSC hg19) using Bowtie⁴¹, allowing three mismatches in the first 28 bases per read (-n3 option). All duplicate reads and those without unique alignment were removed from further analysis. Sequencing, read and mapping data are summarized in Supplementary Table 5. We further analysed only uniquely aligned reads. Each aligned read was extended to a predicted fragment length of 150 bp. Reads were summed in 10-bp bins along the chromosome for ChIP and WCE. To facilitate comparison of detected peaks between different ChIP experiments, we normalized the read number of each bin per million uniquely mapped reads for that chromosome for both ChIP and WCE samples (that is, read number for each bin \times 1,000,000/total number of reads mapped onto reference genomes). This value was further smoothed with a 500 bp width (50 bins). Scanning the genome with a 300 bp (30 bin) sliding window, a one-sided Wilcoxon rank-sum test was performed to estimate the enrichment *P* value for each window. The fold enrichment (ChIP/WCE) for each window was also calculated. To call peaks, we used stringent criteria to identify windows that satisfied both fold enrichment >3.0 and *P* < 1 \times 10⁻⁴ criteria for candidate-binding sites⁴². To eliminate uncertain sites further, we removed the regions with low ChIP reads using the criteria: (1) the average number of ChIP reads in region/the average number of ChIP reads in the genome <3.0; and (2) the maximum read intensity in ChIP bins was less than 0.2 per million reads mapped. Peak calling summary statistics are summarized in Supplementary Table 6.

RNA-Seq analyses. We prepared two replicates for each sample for RNA-sequencing (Solexa HiSeq2000, Illumina). Sequenced RNA reads were aligned to RefSeq RNA database (NM accession numbers only) using Bowtie⁴¹ allowing three mismatches in the first 28 bases per read (-n3 option), and aligned transcript reads were merged for a single gene. To accommodate transcriptional variants, we used both uniquely and multiply aligned reads. Multiply aligned reads were divided equally among all locations (*N*-times matched reads were weighted as 1/*N* reads). Reads that aligned to reference database more than 10 times were discarded. The expression level of each gene was calculated in reads per kilobase per million reads (RPKM)⁴³ and normalized by trimmed mean of *M* values (TMM)⁴⁴. RPKM scores of the two replicates were averaged to calculate the enrichment ratio. Using a 2.5 RPKM threshold, we obtained a total of 9,763 expressed genes (Supplementary Table 7).

30. Hauf, S., Waizenegger, I. C. & Peters, J. M. Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. *Science* **293**, 1320–1323 (2001).
31. Watrin, E. *et al.* Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. *Curr. Biol.* **16**, 863–874 (2006).
32. Eddy, S. R. Multiple alignment using hidden Markov models. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **3**, 114–120 (1995).
33. Wang, Y., Geer, L. Y., Chappey, C., Kans, J. A. & Bryant, S. H. Con3D: sequence and structure views for Entrez. *Trends Biochem. Sci.* **25**, 300–302 (2000).
34. The PyMOL Molecular Graphics System v. 1.5.0.1 (Schrödinger, LLC, 2012).
35. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol.* **26**, 317–325 (2008).
36. Gantt, S. L., Gattis, S. G. & Fierke, C. A. Catalytic activity and inhibition of human histone deacetylase 8 is dependent on the identity of the active site metal ion. *Biochemistry* **45**, 6170–6178 (2006).
37. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
38. Allen, R. C., Zoghbi, H. Y., Moseley, A. B., Rosenblatt, H. M. & Belmont, J. W. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.* **51**, 1229–1239 (1992).
39. Wang, X. *et al.* Mutations in X-linked PORCN, a putative regulator of Wnt signaling, cause focal dermal hypoplasia. *Nature Genet.* **39**, 836–838 (2007).
40. Amos-Landgraf, J. M. *et al.* X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am. J. Hum. Genet.* **79**, 493–499 (2006).
41. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
42. Katou, Y. *et al.* S-phase checkpoint proteins Tof1 and Mrc1 form a stable replication-pausing complex. *Nature* **424**, 1078–1083 (2003).
43. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
44. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).

Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts

Yuji Shiba^{1,2*}, Sarah Fernandes^{1*}, Wei-Zhong Zhu¹, Dominic Filice^{1,3}, Veronica Muskheli¹, Jonathan Kim¹, Nathan J. Palpant¹, Jay Gantz^{1,3}, Kara White Moyes¹, Hans Reinecke¹, Benjamin Van Biber¹, Todd Dardas⁴, John L. Mignone⁴, Atsushi Izawa², Ramy Hanna⁴, Mohan Viswanathan⁴, Joseph D. Gold⁵, Michael I. Kotlikoff⁶, Narine Sarvazyan⁷, Matthew W. Kay^{7,8}, Charles E. Murry^{1,3,4} & Michael A. Laflamme¹

Transplantation studies in mice and rats have shown that human embryonic-stem-cell-derived cardiomyocytes (hESC-CMs) can improve the function of infarcted hearts^{1–3}, but two critical issues related to their electrophysiological behaviour *in vivo* remain unresolved. First, the risk of arrhythmias following hESC-CM transplantation in injured hearts has not been determined. Second, the electromechanical integration of hESC-CMs in injured hearts has not been demonstrated, so it is unclear whether these cells improve contractile function directly through addition of new force-generating units. Here we use a guinea-pig model to show that hESC-CM grafts in injured hearts protect against arrhythmias and can contract synchronously with host muscle. Injured hearts with hESC-CM grafts show improved mechanical function and a significantly reduced incidence of both spontaneous and induced ventricular tachycardia. To assess the activity of hESC-CM grafts *in vivo*, we transplanted hESC-CMs expressing the genetically encoded calcium sensor, GCaMP3 (refs 4, 5). By correlating the GCaMP3 fluorescent signal with the host ECG, we found that grafts in uninjured hearts have consistent 1:1 host-graft coupling. Grafts in injured hearts are more heterogeneous and typically include both coupled and uncoupled regions. Thus, human myocardial grafts meet physiological criteria for true heart regeneration, providing support for the continued development of hESC-based cardiac therapies for both mechanical and electrical repair.

Although hESC-CMs form gap junctions and beat synchronously *in vitro*^{6,7}, there is only indirect evidence for their electromechanical integration after transplantation^{6–8}. We do not know whether hESC-CM grafts contract synchronously at physiological human rates, integrate in injured hearts despite scar tissue, or affect electrical stability. Indeed, both pro-arrhythmic⁹ and anti-arrhythmic¹⁰ effects have been reported for mouse cardiomyocyte grafts in injured mouse hearts. Human cardiac grafts could plausibly contribute to arrhythmogenesis through automaticity^{11–14} and triggered activity¹⁵, and their irregular graft geometry could promote reentry^{1–3,16}.

To address these uncertainties, we employed a new guinea-pig model of cardiac injury. Prior work with hESC-CMs in infarcted hearts was carried out in mice and rats^{1–3}, but these species' rapid heart rates (approximately 600 and 400 beats per minute (b.p.m.), respectively¹⁷) may prevent host-graft coupling or arrhythmias that could occur in humans. *In vitro* hESC-CMs show a spontaneous rate of approximately 50 to 150 b.p.m.^{12,13,18} and can be paced up to 240 b.p.m., suggesting that they can keep up with the guinea-pig heart (approximately 200 to 250 b.p.m.¹⁹).

We first examined the structural, mechanical and electrocardiographic consequences of hESC-CM transplantation in injured hearts of immunosuppressed guinea-pigs (Supplementary Fig. 1a). hESC-CMs were derived from H7 hESCs, as previously described^{2,20}. Adult guinea-pigs were subjected to cardiac cryoinjury and implanted with telemetric electrocardiographic (ECG) transmitters. Ten days later, they underwent a repeat thoracotomy and intra-cardiac injection of either 1×10^8 hESC-CMs in a pro-survival cocktail (PSC) of factors previously shown to enhance hESC-CM engraftment² ($n = 15$), 1×10^8 non-cardiac hESC-derivatives in PSC (non-CMs; $n = 13$), or PSC vehicle alone ($n = 14$). hESC-CMs were 63% pure by anti- α -actinin flow cytometry, whereas non-CMs included no detectable cardiomyocytes (Supplementary Fig. 2).

Twenty-eight days after transplantation, all animals showed transmural scar and thinning of the left ventricle. Scar area was not different among the groups ($13.2 \pm 0.9\%$ of the left ventricle in hESC-CM, $14.8 \pm 1.4\%$ in non-CM, and $15.3 \pm 1.9\%$ in PSC-only recipients). However, hESC-CM recipients showed partial remuscularization with islands of human myocardium occupying $8.4 \pm 1.5\%$ of the scar area (Fig. 1a). The human origin of these grafts was confirmed by *in situ* hybridization with a human pan-centromeric probe, and more than 99% of the human cells immunostained positively with the cardiac marker β -myosin heavy chain (β MHC; also known as MYH7) (Fig. 1b and Supplementary Fig. 3a–f). No teratomas developed, and hESC-CM grafts were negative for multiple non-cardiac markers (Supplementary Fig. 3g–k). Most of the graft myocardium was located in the central scar, but there were occasional points of host-graft contact in the border zone with shared intercalated discs identified by anti-connexin 43 and cadherin immunostaining (Fig. 1c–h and Supplementary Fig. 3d, e). Minimal immune reaction was observed in sections stained with a guinea-pig-specific pan-leukocyte marker (Supplementary Fig. 3l). Grafts were supplied by host-derived neovessels that contained erythrocytes, indicating perfusion by the host coronary circulation (Supplementary Fig. 3h, m–o).

Surviving human cells were found in only 7 of 13 non-CM recipients at 28 days after transplantation, and these grafts were smaller than those in hESC-CM recipients (less than 1% of the scar area). No β MHC-positive graft was detected in non-CM recipients; instead, the surviving human cells consisted of small epithelial nests and scattered fibroblastic cells (Supplementary Fig. 4).

Hearts receiving hESC-CMs, non-CMs and PSC only, were assessed by echocardiography on days –2 and +28 relative to cell transplantation. All groups showed increased left-ventricle dimensions and reduced fractional shortening on day –2 relative to uninjured controls

¹Department of Pathology, Center for Cardiovascular Biology, Institute for Stem Cell and Regenerative Medicine, University of Washington, 850 Republican Street, Seattle, Washington 98109, USA.

²Department of Cardiovascular Medicine, Shinshu University, 3-1-1 Asahi, Matsumoto, Nagano 390-8621, Japan. ³Department of Bioengineering, Center for Cardiovascular Biology, Institute for Stem Cell and Regenerative Medicine, University of Washington, 850 Republican Street, Seattle, Washington 98109, USA. ⁴Department of Medicine and Cardiology, Center for Cardiovascular Biology, Institute for Stem Cell and Regenerative Medicine, University of Washington, 1959 NE Pacific Street, Seattle, Washington 98195, USA. ⁵Geron Corporation, 230 Constitution Drive, Menlo Park, California 94025 USA.

⁶Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA. ⁷Department of Pharmacology and Physiology, The George Washington University, 2300 I Street NW, Washington DC 20037 USA. ⁸Department of Electrical and Computer Engineering, The George Washington University, 2300 I Street NW, Washington DC 20037 USA.

*These authors contributed equally to this work.

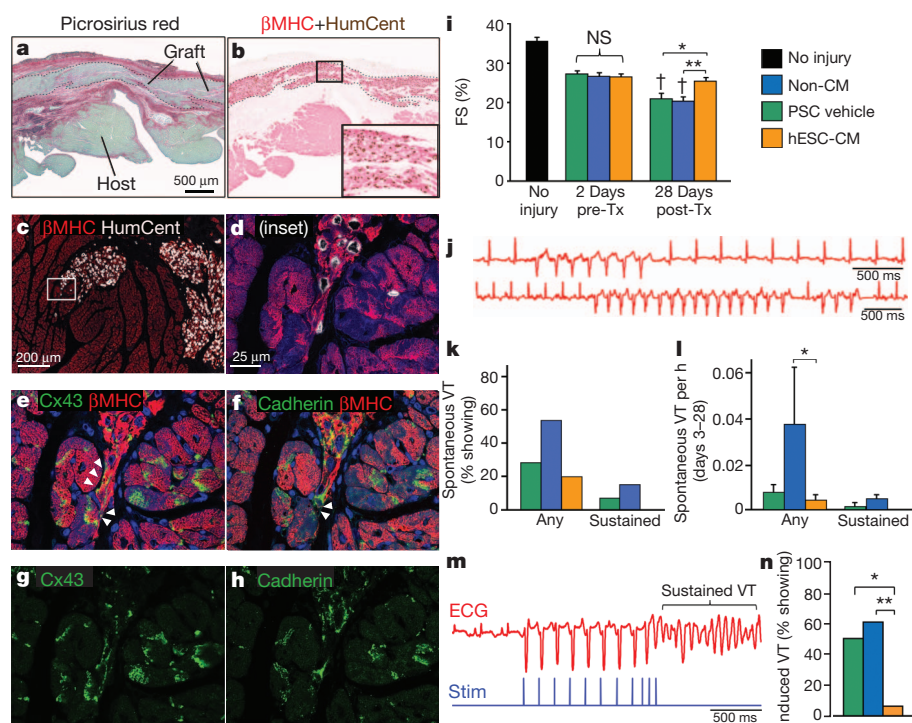


Figure 1 | Transplanted hESC-CMs partially remuscularize injured guinea-pig hearts, preserve mechanical function and reduce arrhythmia susceptibility. **a, b**, Twenty-eight-day-old hESC-CM grafts in a cryoinjured heart stained with picrosirius red (**a**), and anti- β -myosin heavy chain (β MHC, red) plus human-specific *in situ* probe (HumCent, brown) (**b**). **c–h**, Confocal image of host-graft contact, dual-labelled for HumCent (white) and β MHC (red) (**c**); a higher magnification image of the area in the white box in **c** is shown in **d**), de-stained and then immunostained for β MHC (red) and either connexin 43 (Cx43, green; **e** and **g**) or cadherins (green; **f** and **h**). Arrows indicate Cx43 and cadherin shared between graft and host myocytes. **i**, Left-ventricle fractional shortening (FS) by echocardiography in uninjured animals and cryoinjured recipients of PSC vehicle only, non-CMs or hESC-CMs at 2 days prior to and 28 days after transplantation (Tx). **j**, Representative telemetric ECGs, including non-sustained ventricular tachycardia (VT) in an hESC-CM recipient (top trace), as well as sustained VT and triplet PVCs in a non-CM recipient (bottom trace). **k**, Percentage of animals by group that showed spontaneous VT during monitoring from days 3 to 28 after transplantation. **l**, Frequency of spontaneous VT by group. **m**, ECG (red) and stimulation (blue) traces from a cryoinjured non-CM recipient induced to sustained VT by PES. **n**, Percentage of animals in each group induced to VT. All data are presented as mean \pm s.e.m.; $n \geq 13$ per group. NS, not significant. * $P < 0.05$; ** $P < 0.01$. † $P < 0.05$ versus day -2 .

($n = 11$), but there were no differences among the cryoinjured groups (Fig. 1i and Supplementary Fig. 5). The two control groups exhibited further left ventricle dilation and deterioration of fractional shortening between days -2 and $+28$ ($P < 0.05$ for all pairwise comparisons between time points). This deterioration was completely attenuated in cryoinjured hESC-CM recipients, which showed significantly greater fractional shortening on day $+28$ than cryoinjured PSC-only or non-CM recipients ($P < 0.05$ and $P < 0.01$, respectively) (Fig. 1i).

Ambulatory telemetric ECG recordings were obtained regularly from 1 day after injury to 28 days after transplantation. Pilot studies revealed only rare premature ventricular contractions (PVCs) in uninjured animals, whereas cryoinjured but not transplanted animals showed occasional PVCs and runs of ventricular tachycardia (Supplementary Table 1). Interestingly, the hESC-CM group had the lowest fraction of animals with PVCs after transplantation and the lowest rate of couplet PVCs per hour (Supplementary Fig. 6). Moreover, fewer hESC-CM recipients showed spontaneous ventricular tachycardia, and hESC-CM recipients were the only group with no sustained ventricular tachycardia (Fig. 1j, k). Next, we counted all episodes of ventricular tachycardia from day 3 to 28 and found that non-CM recipients showed 785% more ventricular tachycardia episodes than hESC-CM recipients (95% confidence interval: 74% to $+4,370\%$, $P < 0.01$; Fig. 1l).

As a final test of electrical stability, we induced arrhythmias through programmed electrical stimulation (PES). To detect either a positive or negative effect of cell transplantation, we developed a PES protocol that induced ventricular tachycardia in 40% of non-transplanted animals at 38 days after cryoinjury ($n = 5$), but not in uninjured animals ($n = 8$). When cryoinjured hESC-CM recipients were challenged with PES at 28 days after transplantation, ventricular tachycardia was induced in only 6.7% of these animals versus 61.5% and 50% of cryoinjured non-CM and PSC-only recipients, respectively ($P < 0.01$ by Fisher's exact test, Fig. 1m, n).

To explore the mechanism by which hESC-CMs improved mechanical and electrical function, we investigated whether the grafts were

coupled and beat synchronously with host myocardium. We used zinc-finger nuclease (ZFN)-mediated transgenesis²¹ to create hESC-CMs that stably expressed the genetically encoded fluorescent calcium sensor, GCaMP3 (refs 4, 5), and these myocytes exhibited robust fluorescence transients with each contractile cycle *in vitro* (Supplementary Fig. 7 and Supplementary Movie 1). We transplanted 1×10^8 GCaMP3-positive hESC-CMs (cardiac purity of $74 \pm 10\%$) into intact and cryoinjured guinea-pig hearts and carried out intravital imaging at 2 or 4 weeks after transplantation (Supplementary Fig. 1b). Epicardial fluorescent transients indicated activation of the graft *in vivo*, and these were correlated with the ECG to determine their synchrony with host myocardium¹⁰.

The GCaMP3-positive hESC-CM grafts in uninjured hearts had extensive host-graft contact with little intervening fibrosis (Fig. 2a and Supplementary Fig. 8a). Correspondingly, these hearts showed an impressive degree of host-graft electromechanical integration. We found large GCaMP3-positive hESC-CMs grafts in all uninjured recipients at 14 days after transplantation ($n = 9$), and 100% of the visible graft in each animal showed calcium fluorescence transients that synchronized 1:1 with systole in the host ECG (Fig. 2b). These hearts were also imaged *ex vivo* during mechanical arrest with either 2,3-butanedione monoxime (BDM) or blebbistatin ($n = 11$). 1:1 host-graft coupling continued under these conditions, ruling out confounding motion artefacts and the possibility of indirect graft activation by passive stretch (Supplementary Fig. 8 b, c and Supplementary Movie 2). hESC-CM grafts in uninjured hearts remained fully coupled during pacing at rates ≥ 5 Hz (Fig. 2c).

Next, we transplanted GCaMP3-positive hESC-CMs at 10 days after cryoinjury and carried out intravital imaging at 14 and 28 days after transplantation ($n = 7$ and $n = 15$ animals, respectively). By histology, surviving grafts at both time points were mostly located within the central scar, although occasional points of host-graft contact in the border zone were identified (Fig. 3a and Supplementary Fig. 9a). Dual labelling with a guinea-pig-specific *in situ* probe and anti-green fluorescent protein (GFP) immunostaining ruled out a meaningful

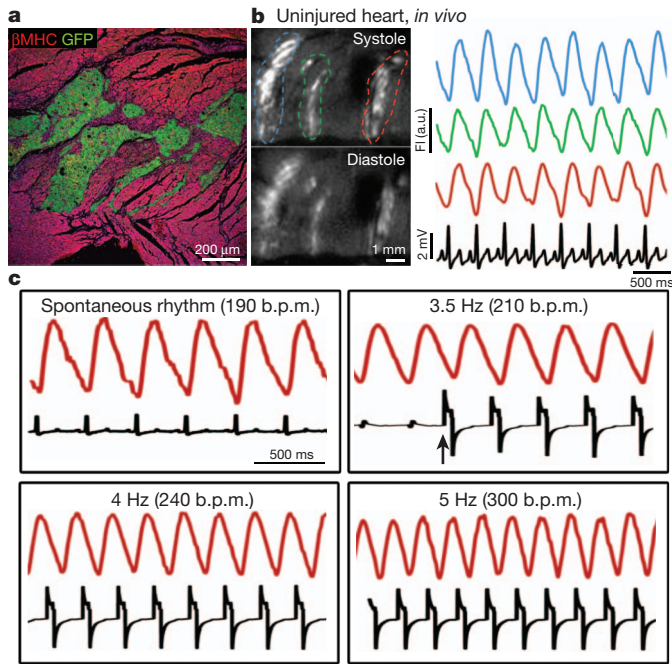


Figure 2 | hESC-CM grafts in uninjured hearts show 1:1 coupling with host myocardium. **a**, GCaMP3-positive hESC-CM graft in an uninjured heart immunostained for GFP (green) and β MHC (red), showing extensive host-graft contact with minimal intervening scar. **b**, Left, GCaMP3-positive hESC-CM graft in an uninjured heart during systole and diastole, acquired using an open-chest preparation at 14 days after transplantation. Right, GCaMP3 fluorescence intensity (FI) versus time for the red, green and blue regions of interest indicated in the left panel, as well as the host ECG (black). GCaMP3 fluorescent transients in all three regions occurred with 1:1 coupling with QRS complexes of the host ECG. a.u. arbitrary units. **c**, GCaMP3 fluorescent signal (red) and ECG (black) from a representative hESC-CM graft in an uninjured heart, imaged *ex vivo* under mechanical arrest. 1:1 host-graft coupling occurred during spontaneous beating and pacing at rates ≤ 5 Hz. Arrow, onset of pacing.

contribution of cell fusion to the observed GCaMP3 signals (Supplementary Fig. 9b and Supplementary Movie 3).

Cryoinjured hearts with GCaMP3-positive hESC-CM grafts showed robust epicardial fluorescence transients, but their behaviour was more heterogeneous than that of grafts in uninjured animals, with greater variation in size, distribution and temporal relationship to the host ECG. Given this complicated behaviour, we focused on imaging mechanically arrested hearts *ex vivo*, so that the grafts could be more completely visualized and the proportion of coupled and uncoupled graft could be quantified better (Supplementary Movies 4 and 5). We found regions of GCaMP3-positive graft with consistent 1:1 host-graft coupling in approximately 60% of hearts at both 14 and 28 days after transplantation (4 out of 7 animals, and 9 out of 15 animals, respectively; Fig. 3b and Supplementary Tables 2 and 3). Although there was no significant difference in the fraction of animals with coupled grafts between the two time points, in the subset of animals with at least some coupling, there was a trend towards greater host-graft coupling over time ($66 \pm 10\%$ versus $87 \pm 6\%$ of visible graft at 14 and 28 days, respectively). No grafts with intermittent or partial coupling (for example, 1:2 or 1:3 host-graft coupling) were observed. Uncoupled graft regions showed regular GCaMP3 fluorescent transients, but these occurred with a periodicity unrelated to the ECG or neighbouring graft domains, usually at a rate slower than the host (at 28 days, 96 ± 11 b.p.m. versus a host rate of 176 ± 9 b.p.m.). By marking the coupled and non-coupled graft regions with indelible inks that survive histological processing, we found many examples of reliably coupled graft located within the central scar region, proving that coupling is not limited to grafts at the border zone (data not shown). When injured hearts with

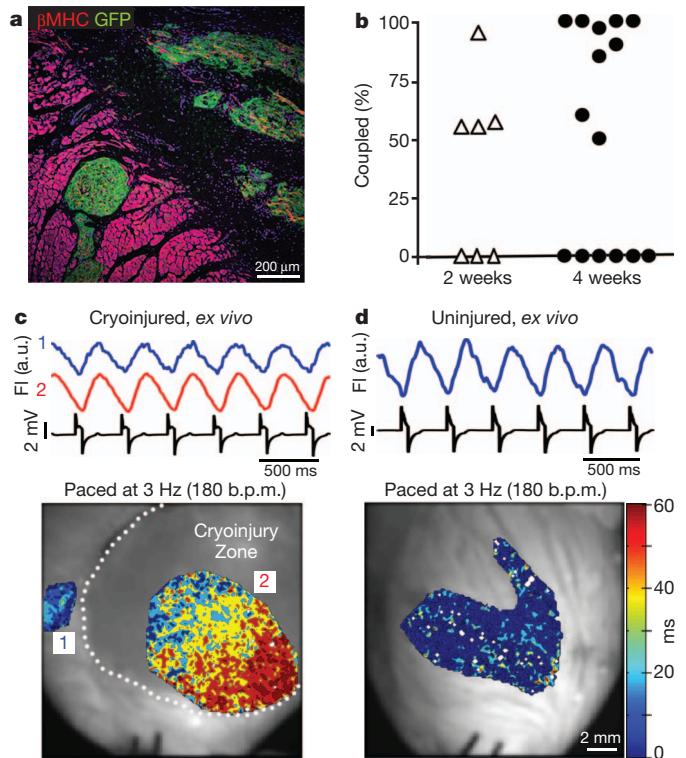


Figure 3 | hESC-CM grafts show 1:1 host-graft coupling in the majority of injured recipient hearts, but the extent of coupling and pattern of activation is variable. **a**, GCaMP3-positive hESC-CM graft in a cryoinjured heart immunostained for GFP (green) and β MHC (red). Small graft nests were located in host muscle within the border zone, but most of the graft was in scar. **b**, Percentage of visible GCaMP3-positive hESC-CM graft in each cryoinjured heart that showed 1:1 host-graft coupling at 14 and 28 days after transplantation ($n = 7$ and $n = 15$ animals, respectively). **c**, Representative 28-day-old GCaMP3-positive hESC-CM graft in a cryoinjured heart imaged *ex vivo* during mechanical arrest with blebbistatin and pacing at 3 Hz. Top, traces of mean fluorescent intensity versus time for graft regions located within host muscle (1, blue font) and the cryoinjury zone (2, red font). Both were activated in a 1:1 correlation with the host ECG (black). Bottom, corresponding activation map showing the interval (in ms) between the stimulus pulse and the local rise in GCaMP3 fluorescence. Graft in host muscle (1) showed uniformly rapid activation, whereas the graft in the scar (2) activated first in centre of the scar and then gradually progressed towards the border zone. On other occasions, graft activation started at the border zone and radiated into the scar. **d**, Top, traces of GCaMP3 fluorescence and ECG. Bottom, the activation map for a representative GCaMP3-positive hESC-CM graft in a blebbistatin-arrested uninjured heart. This graft showed 1:1 host-graft coupling and a brief interval between stimulus and GCaMP3 transient upstroke.

GCaMP3-positive hESC-CM grafts were paced, most of the initially coupled graft regions remained coupled up to 4 Hz, after which some loss of coupling was observed (Supplementary Fig. 10).

We next attempted to perform voltage mapping of graft and host myocardium in cryoinjured hearts but, unexpectedly, hESC-CM grafts failed to label with the potentiometric dyes RH237 and di-4-ANEPPS (Supplementary Fig. 11). hESC-CMs were readily labelled *in vitro* (Supplementary Fig. 12), so their poor staining *in vivo* probably reflects sluggish perfusion of the graft tissue and perhaps reduced membrane staining because of undeveloped T-tubules²². However, we did obtain good labelling of host myocardium with RH237 and we therefore compared voltage maps from cryoinjured hearts with and without grafts ($n = 6$ and $n = 4$ animals, respectively). As expected, both groups showed robust propagation in uninjured myocardium, whereas optical action potential amplitudes and epicardial conduction velocities were reduced in the cryoinjury zone. The two groups showed no significant differences in the amplitudes of optical action potentials within the injured zone or the kinetics of activation and repolarization

(Supplementary Figs 13 and 14), suggesting that hESC-CM transplantation does not exert major paracrine effects on host electrical behaviour.

Because the electrical activity of hESC-CM grafts could not be evaluated with potentiometric dyes, we used instead the GCaMP3 signal to assess graft activation and propagation. GCaMP3 transients lag depolarization owing to slow binding of calcium, but this delay should be the same throughout the graft^{10,23}, so graft activation maps can be generated using the time interval between electrical stimulation (or host QRS complex of the ECG) and the GCaMP3 transient. This revealed a marked difference in activation between grafts located in viable host muscle versus those in scar tissue (Fig. 3c, d). Grafts in intact myocardium were activated uniformly and within approximately 10 ms, probably reflecting synchronous activation through multiple points of contact with the host. In contrast, graft activation within scar tissue was spatially and temporally heterogeneous, even in regions with 1:1 coupling. These coupled grafts showed initial activation within approximately 10 to 20 ms, but the spread of activation was slower, such that some areas required >50 ms for activation. This probably reflects the more limited contacts between graft and host, as well as slower propagation within the immature graft myocardium.

These intravital imaging studies are, to our knowledge, the first direct demonstration that human cardiomyocytes can integrate and contract synchronously with host myocardium. Although additional paracrine mechanisms cannot be excluded, the demonstration of electromechanically coupled grafts in injured hearts supports the idea that hESC-CMs can improve mechanical function by creating new force-generating units, a sine qua non for heart regeneration. However, hESC-CM grafts in injured hearts were not always fully integrated, implying that enhancing their integration may further benefit contractile function. The model developed here represents a reasonably high-throughput platform to test strategies to improve host-graft integration (for example, tissue engineering, attenuating graft cell death and reducing fibrosis).

Our study also provides reassurance about the arrhythmogenic risk of cardiac repair with immature stem-cell-derived cardiomyocytes. Consistent with prior transplantation studies with primary fetal mouse cardiomyocytes¹⁰, we observed an arrhythmia-suppressive effect that was unique to cardiomyocyte grafts and occurred despite incomplete host-graft coupling. Although this is encouraging, before clinical translation of hESC-based therapies, it will be necessary to carry out safety and efficacy studies with longer follow-up durations and in a large animal model.

METHODS SUMMARY

GCaMP3-positive hESCs were generated by ZFN-mediated targeting, following a method described previously²¹ (Supplementary Fig. 7), and these transgenic and wild-type hESCs were differentiated into cardiomyocytes, as previously reported^{2,20}. Guinea-pigs underwent cryoinjury, immunosuppression with cyclosporine plus methylprednisolone, cell transplantation and physiological monitoring as detailed in the Supplementary Methods and Supplementary Fig. 1. Intravital imaging of GCaMP3-positive hESC-CM grafts was carried out on day 14 or 28 after transplantation, using open chest and *ex vivo* Langendorff preparations. GCaMP3 and RH237 fluorescent signals were acquired with a high-speed camera (Andor iXon 860 EM-CCD) on an epifluorescence stereomicroscope. All procedures complied with regulations of the University of Washington Institutional Animal Care and Use Committee.

Full Methods and any associated references are available in the online version of the paper.

Received 26 April 2011; accepted 12 June 2012.

Published online 5 August 2012.

1. Caspi, O. *et al.* Transplantation of human embryonic stem cell-derived cardiomyocytes improves myocardial performance in infarcted rat hearts. *J. Am. Coll. Cardiol.* **50**, 1884–1893 (2007).
2. Laflamme, M. A. *et al.* Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature Biotechnol.* **25**, 1015–1024 (2007).

3. van Laake, L. W. *et al.* Human embryonic stem cell-derived cardiomyocytes survive and mature in the mouse heart and transiently improve function after myocardial infarction. *Stem Cell Res.* **1**, 9–24 (2007).
4. Tallini, Y. N. *et al.* Imaging cellular signals in the heart *in vivo*: cardiac expression of the high-signal Ca^{2+} indicator GCaMP2. *Proc. Natl Acad. Sci. USA* **103**, 4753–4758 (2006).
5. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
6. Kehat, I. *et al.* Electromechanical integration of cardiomyocytes derived from human embryonic stem cells. *Nature Biotechnol.* **22**, 1282–1289 (2004).
7. Xue, T. *et al.* Functional integration of electrically active cardiac derivatives from genetically engineered human embryonic stem cells with quiescent recipient ventricular cardiomyocytes: insights into the development of cell-based pacemakers. *Circulation* **111**, 11–20 (2005).
8. Gepstein, L. *et al.* *In vivo* assessment of the electrophysiological integration and arrhythmogenic risk of myocardial cell transplantation strategies. *Stem Cells* **28**, 2151–2161 (2010); erratum **29**, 1475 (2011).
9. Liao, S. Y. *et al.* Proarrhythmic risk of embryonic stem cell-derived cardiomyocyte transplantation in infarcted myocardium. *Heart Rhythm* **7**, 1852–1859 (2010).
10. Roell, W. *et al.* Engraftment of connexin 43-expressing cells prevents post-infarct arrhythmia. *Nature* **450**, 819–824 (2007).
11. Kehat, I. *et al.* Human embryonic stem cells can differentiate into myocytes with structural and functional properties of cardiomyocytes. *J. Clin. Invest.* **108**, 407–414 (2001).
12. He, J. Q., Ma, Y., Lee, Y., Thomson, J. A. & Kamp, T. J. Human embryonic stem cells develop into multiple types of cardiac myocytes: action potential characterization. *Circ. Res.* **93**, 32–39 (2003).
13. Zhu, W. Z. *et al.* Neuregulin/ErbB signaling regulates cardiac subtype specification in differentiating human embryonic stem cells. *Circ. Res.* **107**, 776–786 (2010).
14. Mandel, Y. *et al.* Human embryonic and induced pluripotent stem cell-derived cardiomyocytes exhibit beat rate variability and power-law behavior. *Circulation* **125**, 883–893 (2012).
15. Jonsson, M. K. *et al.* Quantified proarrhythmic potential of selected human embryonic stem cell-derived cardiomyocytes. *Stem Cell Res.* **4**, 189–200 (2010).
16. Chen, H. S., Kim, C. & Mercola, M. Electrophysiological challenges of cell-based myocardial repair. *Circulation* **120**, 2496–2508 (2009).
17. Swoap, S. J., Overton, J. M. & Garber, G. Effect of ambient temperature on cardiovascular parameters in rats and mice: a comparative approach. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287**, R391–R396 (2004).
18. Mummery, C. *et al.* Differentiation of human embryonic stem cells to cardiomyocytes: role of coculture with visceral endoderm-like cells. *Circulation* **107**, 2733–2740 (2003).
19. Shiotani, M., Harada, T., Abe, J., Hamada, Y. & Horii, I. Methodological validation of an existing telemetry system for QT evaluation in conscious guinea pigs. *J. Pharmacol. Toxicol. Methods* **55**, 27–34 (2007).
20. Zhu, W. Z., Van Biber, B. & Laflamme, M. A. Methods for the derivation and use of cardiomyocytes from human pluripotent stem cells. *Methods Mol. Biol.* **767**, 419–431 (2011).
21. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnol.* **27**, 851–857 (2009).
22. Lieu, D. K. *et al.* Absence of transverse tubules contributes to non-uniform Ca^{2+} wavefronts in mouse and human embryonic stem cell-derived cardiomyocytes. *Stem Cells Dev.* **18**, 1493–1500 (2009).
23. Kotlikoff, M. I. Genetically encoded Ca^{2+} indicators: using genetics and molecular design to understand complex physiology. *J. Physiol. (Lond.)* **578**, 55–67 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Tallini, L. Linares, B. Johnson and S. Dupras for advice and technical assistance. This work was partly supported by a grant from Geron Corporation (M.A.L.), as well as by US National Institutes of Health grants K08-HL80431 (M.A.L.), R01-HL064387 (M.A.L. and C.E.M.), P01-HL094374 (M.A.L. and C.E.M.), R01-HL084642 (C.E.M.), P01-GM81619 (C.E.M.), U01-HL100405 (M.A.L. and C.E.M.) and R01-HL095828 (N.S. and M.W.K.). Animal experiments were supported in part by the University of Washington's Mouse Metabolic Phenotyping Center, U24-DK076126.

Author Contributions S.F., Y.S., W.-Z.Z., D.F., M.I.K., J.D.G., C.E.M. and M.A.L. designed the study. Y.S. and S.F. led the arrhythmia and calcium imaging experiments, respectively. Y.S. developed and performed the telemetry and programmed electrical stimulation. T.D., J.L.M., A.I., R.H. and M.V. performed telemetric ECG interpretation. The GCaMP3-expressing hESC line was generated by J.G., N.J.P. and B.V.B.V.M., J.K., K.W.M., S.F. and Y.S. carried out and analysed immunohistochemistry experiments. S.F., H.R. and M.A.L. developed and performed guinea-pig-specific *in situ* hybridization. S.F., W.-Z.Z. and D.F. carried out and analysed the GCaMP3 imaging experiments. D.F., M.W.K. and N.S. developed and analysed the voltage mapping experiments. All authors contributed to data analysis and interpretation. Y.S. created the figures with the assistance of S.F., W.-Z.Z., K.W.M. and D.F. C.E.M. and M.A.L. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.A.L. (laflamme@u.washington.edu) or C.E.M. (munry@uw.edu).

METHODS

Cell preparation. Undifferentiated H7 hESCs²⁴ were expanded using either mouse embryonic fibroblast-conditioned medium (MEF-CM)²⁵ or a defined medium supplemented with basic fibroblast growth factor and transforming growth factor β 1 (R&D Systems)²⁶. hESCs were then differentiated into cardiomyocytes using our previously reported directed differentiation protocol that involves the serial application of activin A and bone morphogenetic protein 4 (BMP4, R&D) under defined, serum-free, monolayer culture conditions^{2,13,20}. Non-cardiac control cells were generated by subjecting hESCs to the same protocol but without the addition of activin A and BMP4. See Supplementary Fig. 2 for additional immunophenotypic information regarding the two cell preparations employed.

Human ESC-CMs and non-cardiac hESC derivatives were collected and cryopreserved after 16 to 18 days under differentiating conditions. One day before collection, cells were subjected to a pro-survival protocol, previously shown to enhance engraftment after transplantation². In brief, cultures were heat-shocked with a 30-min exposure to 43 °C medium, followed by RPMI-B27 medium supplemented with IGF1 (100 ng ml⁻¹, Peprotech) and cyclosporine A (0.2 μ M, Sandimmune, Novartis). One day later, cultures were collected with 0.25% trypsin per 0.5 mM EDTA (Invitrogen) and cryopreserved as described previously²⁷.

Immediately before transplantation, cells were thawed at 37 °C, washed with RPMI, and suspended in a 150- μ l volume (per animal) of pro-survival cocktail (PSC)², which consisted of 50% (v/v) growth factor-reduced Matrigel, supplemented with ZVAD (100 mM, benzoyloxycarbonyl-Val-Ala-Asp(O-methyl)-fluoromethyl ketone, Calbiochem), Bcl-XL BH4 (cell-permeant TAT peptide, 50 nM, Calbiochem), cyclosporine A (200 nM, Wako), IGF1 (100 ng ml⁻¹, Peprotech) and pinacidil (50 mM, Sigma).

Surgical procedures. For cardiac cryoinjury and ECG transmitter placement, male guinea-pigs (650–750 g, Charles River) were anaesthetized with an intraperitoneal injection of 50 mg kg⁻¹ ketamine and 2 mg kg⁻¹ xylazine. The skin was shaved and sterilized, a small incision was made in the right flank and the transmitter (PhysioTel model CA-F40, DSI) was inserted in a subcutaneous pocket. The positive and negative leads were then tunnelled subcutaneously to the left fifth intercostal space and the upper sternal midline, respectively. The animal was then intubated, mechanically ventilated and anaesthetized with 1.5% isoflurane. The heart was exposed by a left thoracotomy, the pericardium was opened and a 10-mm-diameter aluminium cryoprobe, pre-cooled with liquid nitrogen, was applied to the left ventricular free wall 4 times for 30 s each. At the end of the surgery, all animals received 0.025% (v/v) topical bupivacaine at the wound site, as well as 0.05 mg kg⁻¹ intraperitoneal buprenorphine for analgesia.

From day -2 to day +28 relative to cell transplantation, animals were treated with an immunosuppressive regimen of methylprednisolone (2 mg kg⁻¹ per day, intraperitoneal) and cyclosporine A (15 mg kg⁻¹ per day, subcutaneous, for 7 days; thereafter reduced to 7.5 mg kg⁻¹ per day). The trough level of blood cyclosporine was measured on day +28 at $838 \pm 64 \mu$ g l⁻¹. Cell transplantation was carried out using a repeat thoracotomy and injecting the total bolus of cells into three separate injection sites, that is, the central cryolesion and the flanking lateral border zones.

Echocardiography. On days -2 and +28 relative to cell transplantation, animals were lightly anaesthetized with inhaled 1.5% isoflurane, and then their left-ventricular end-diastolic dimension (LVEDD), left-ventricular end-systolic dimension (LVESD) and heart rate were measured by transthoracic echocardiography (GE Vivid 7) with a 10-MHz paediatric transducer. Fractional shortening (FS) was calculated using this equation: $FS = 100 \times ((LVEDD - LVESD) / LVEDD)$. All echocardiographic scans and analyses were performed by an operator who was blind to the experimental conditions.

Telemetric electrocardiography. ECG recordings were acquired from conscious, freely mobile animals using a Dataquest ART telemetry system (DSI)^{19,28}. All recordings were obtained at the same time of day (evening) for 6 h (3 h in the light, 3 h in the dark) during a period in which the animals' living quarters were undisturbed. Recordings were obtained from cryoinjured recipients of hESC-CMs, non-CMs or PSC only, on days 1, 6 and 9 after injury and days 1, 3, 7, 10, 14, 17, 21, 24 and 27 after transplantation. All ECG traces were evaluated manually by a cardiologist who was blind to the experimental conditions. The cardiologist determined the total number and frequency of events including single and multiform PVCs, as well as non-sustained and sustained ventricular tachycardia. In accordance with Lambeth convention guidelines²⁹, ventricular tachycardia was defined as a run of 4 or more PVCs, and sustained ventricular tachycardia was defined as a fast ventricular rhythm of more than 15 beats.

Programmed electrical stimulation. PES studies were performed on day +28 after transplantation, using methods modified from a previous study³⁰. In brief, each animal was mechanically ventilated, anaesthetized with 2% isoflurane, and outfitted for standard surface ECG recordings (ADInstruments). A midline incision was made in the epigastric region, and a custom-designed bipolar stimulating electrode (FHC) was inserted through the diaphragm into direct contact

with the cardiac apex. PES studies were then carried out using a stimulator generator (STG-1000, MultiChannel Systems) with the pulse output set at twice the capture threshold and the pulse width at 1 ms. We used standard clinical PES protocols, including the application of single, double and triple extra stimuli after a train of 8 conditioning stimuli at a 150-ms cycle length. To determine the ventricular effective refractory period (VERP), a first extra stimulus (S2) was applied with the S1–S2 interval decreased in 5-ms increments from 150 ms until the occurrence of loss of capture. The heart was then challenged three times with the train of eight followed by the single extra stimulus (with the S1–S2 interval set at VERP + 10 ms). If no ventricular tachycardia was induced, this procedure was repeated to apply three challenges with double and, if necessary, triple extra stimuli. In each case, the coupling interval between the final two pulses was set at the VERP for the last extra stimulus + 10 ms. Ventricular tachycardia was only induced in cryoinjured animals with the application of a triple extra stimulus. Notably, ventricular-tachycardia induction by PES was strongly associated with a history of spontaneous ventricular tachycardia during the final day of telemetry monitoring ($P < 0.001$, McNemar's test).

Generation of the GCaMP reporter hESC line. A transgene encoding for the constitutive expression of GCaMP3 was inserted into the AAVS1 locus in H7 hESCs, using methods adapted from a previous study²¹ (see Supplementary Fig. 7). In brief, the right and left arms of an AAVS1-specific ZFN were *de novo* synthesized (Genscript) and cloned into a single polycistronic plasmid in which the expression of each was driven by an independent human PGK promoter. A second polycistronic vector was generated in which (approximately 800-bp) homology arms flanking the AAVS1 ZFN cut site (pZDonor, Sigma Aldrich) surrounded a 5.1-kb insert with two elements: a cassette in which the CAG promoter drives expression of GCaMP3 (ref. 5) (Addgene, plasmid #22692) and a second cassette encoding for PGK-driven expression of neomycin resistance.

AAVS1 ZFN (5- μ g) and AAVS1 CAG GCaMP3 targeting vector (40- μ g) plasmids were co-electroporated (Lonza, Nucleofection system) into H7 hESCs cultured in MEF-CM supplemented with 10 μ M Y-27632. Green fluorescent colonies were isolated and expanded and selected with 40 to 100 μ g ml⁻¹ G418 (Invitrogen) for 5 to 10 days. GCaMP3 H7 hESCs showed a normal karyotype.

Southern blot analysis. Wild-type and transgenic GCaMP3-positive hESC genomic DNA were digested with the restriction enzyme SphI, run on 1% polyacrylamide gel and transferred to a membrane (BioRad Zeta Probe). The membrane was washed in 2 \times SSC and dried at 80 °C in a hybridization oven for 2 h, followed by 1 h of pre-hybridization in 50% formamide, 0.12 M NaH₂PO₄, 0.25 M NaCl, 7% SDS, and 1 mM EDTA at 43 °C. A hybridization probe was generated with the following primers: CCTGTTAGGCAGATTCCTTATC (sense), AGATGGTGGACGAGGAAGGGG (antisense). The probe was labelled with ³²P dCTP (Amersham Megaprime DNA labelling system) and hybridized overnight in hybridization buffer at 43 °C. After 24 h, the membrane was washed for 20 min with 2 \times SSC/0.1% SDS followed by 20 min in 0.1 \times SSC/0.1% SDS. The membrane was then exposed to autoradiographic film for 3 days. The wild-type band is expected at 6.5 kb, whereas the targeted locus shifts to 2.9 kb.

Imaging of GCaMP3-expressing grafts. Intravital imaging of hearts with GCaMP3-positive grafts was performed on days 14 and 28 after transplantation using either an open-chest or *ex vivo* preparation. For the open-chest preparation, guinea-pigs were anaesthetized, mechanically ventilated and outfitted for standard surface ECG recordings, as described above. The anterior epicardium was then exposed by a wide thoracotomy and visualized using an epifluorescence stereomicroscope (Nikon, SMZ 1000) equipped with an EXFO X-Cite illumination source. GCaMP3 was excited at 450 to 490 nm and bandpass filtered (500 to 550 nm) before detection by an electron-multiplying, charge-coupled device camera (Andor iXon 860 EM-CCD) controlled by Andor Solis software. GCaMP3 image acquisition was typically at 80 to 140 frames per second (f.p.s.). Signals from the charge-coupled device (CCD) camera and the surface ECG were fed through a computer for digital storage and off-line analysis using Andor software and custom Matlab scripts (MathWorks).

Some engrafted hearts were also imaged *ex vivo* after mechanical arrest to eliminate motion artefacts. For these experiments, the heart was collected, rapidly mounted on a gravity-fed Langendorff apparatus and then perfused with modified Tyrode solution at 37 °C. The epicardial GCaMP3 signal was then recorded before and after supplementation of the perfusate with an excitation-contraction uncoupler, either 2,3-butanedione monoxime (BDM, 20 mM)^{31,32} or blebbistatin (10 μ M)³³. The utility of blebbistatin was limited by its substantial blue-green fluorescence³⁴, which interfered with the acquisition of the GCaMP3 signal, so BDM was used in experiments except where otherwise indicated (see Supplementary Table 3). All quantitative comparisons were carried out using hearts treated with the same uncoupler. The electrical activity of the isolated heart was continuously monitored by placing positive and negative ECG leads at the base of the right ventricle and left ventricular apex, respectively. Although most

imaging experiments were carried out under spontaneous conditions, a subset of hearts was also imaged during external pacing at rates from 3 to 6 Hz (ML866 PowerLab 430 Data Acquisition System). Experiments involving pacing at rates of greater than 5 Hz were sometimes limited by the occurrence of irreversible tachyarrhythmias.

In a subset of hearts, GCaMP3 activation maps were generated by determining the interval between a pacing stimulus (or the host QRS complex) and the initial rise in GCaMP3 fluorescence. For this analysis, raw data from the green fluorescence and ECG channels were read into a custom Matlab script for analysis. The GCaMP3 fluorescent signal was median filtered in time to reduce noise, smoothed with a spatial conical filter, and background subtracted. Regions of graft were user defined and then analysed to determine the relative GCaMP3 activation time on a per-pixel basis. Relative GCaMP3 activation times for each pixel were then averaged over several stimuli to generate GCaMP3 activation maps.

Voltage-mapping studies. The preceding imaging system and experimental preparation was also used in optical mapping studies with the potentiometric dye RH237. RH237 was selected as our dye of choice, given our past experience with this indicator and its reasonable spectral separation from GCaMP3. When examined in the same heart, GCaMP3 (green-channel) and RH237 (red-channel) fluorescent signals were acquired consecutively by an exchange of filter cubes, not simultaneously.

In brief, cryoinjured hearts with or without 28-day-old GCaMP3-positive hESC-CM grafts were mounted on a Langendorff apparatus and loaded with RH237 (40 μ M) by a bolus injection (in 10 ml buffer) into the aorta. After an initial stabilization period, RH237-stained hearts were then imaged during both spontaneous and paced rhythms. RH237 was excited at 520 to 550 nm, and emitted light was long-pass filtered (670 nm) before detection at 500 f.p.s. using the high-speed EM-CCD system mentioned above. The resultant fluorescent signals were analysed as described previously^{35–37}. In brief, background fluorescence was subtracted and the signal at each pixel was smoothed using a median temporal filter and a spatial cone filter. The signal at each pixel was then scaled as a percentage of the maximum optical action potential amplitude for all the pixels. Although not simultaneously acquired, RH237 and GCaMP3 signals could be partially correlated in a subset of paced hearts by aligning repetitive signals relative to the stimulating electrode.

Histology. Histological studies were carried out as detailed previously by our group^{2,38,39}. For immunohistochemistry, we used the primary antibodies detailed in Supplementary Table 4, then either fluorescent secondary antibodies (Alexa-conjugated, species-specific antibodies from Molecular Probes) or the avidin-biotin reaction followed by chromogenic detection (ABC kits from Vector Labs). The guinea-pig-specific *in situ* hybridization probe was generated using methods adapted from another study⁴⁰. In brief, guinea-pig genomic DNA was amplified using PCR primers reported previously⁴⁰ (sense: 5'-CTCCTGTCCTGCATCCACT-3'; antisense: 5'-GGATATGAGAGACAGTGGTG-3'). The resultant 345-bp band was excised and then amplified in a second round of PCR doped with digoxigenin-11-dUTP. Subsequent *in situ* hybridization with this digoxigenin-labelled guinea-pig-specific probe was carried out using methods detailed previously for the human-specific pan-centromeric probe². The hybridized probe was detected using a peroxidase-conjugated anti-digoxigenin antibody (Roche), followed either by a chromogenic peroxidase substrate or fluorescent tyramide signal amplification (Molecular Probes).

Statistical analysis. All data were analysed in a blind manner, with the identifier code revealed only after the analysis was completed. All values were expressed as mean \pm standard error. Statistical analyses were performed using PASW Statistics

18, Graphpad Prism and SAS software, with the threshold for significance level set at $P < 0.05$. Echocardiographic outcomes were analysed using an analysis of variance (ANOVA), then by post-hoc comparisons between groups by Tukey HSD, and time-course changes were analysed using paired *t*-test analysis of means. For comparisons of the fraction of animals showing spontaneous or induced arrhythmias, we used a two-sided Fisher's exact test. The incidence of spontaneous arrhythmias was analysed using a Poisson regression model to adjust for repeated measures and the effect of time. This approach was used to calculate the ratio of the observed frequencies of recurrence for the two groups and the 95% confidence interval. We used the McNemar's test for correlated proportions to test for an association between induced ventricular tachycardia and spontaneous ventricular tachycardia either on the final day of monitoring or throughout days 3 to 28 after the transplantation monitoring period. To test for synchrony between GCaMP3 fluorescence and ECG signals, intervals were compared using ANOVA and then a post-hoc Dunnett's test.

24. Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
25. Xu, C. *et al.* Feeder-free growth of undifferentiated human embryonic stem cells. *Nature Biotechnol.* **19**, 971–974 (2001).
26. Li, Y., Powell, S., Brunette, E., Lebkowski, J. & Mandalam, R. Expansion of human embryonic stem cells in defined serum-free medium devoid of animal-derived products. *Biotechnol. Bioeng.* **91**, 688–698 (2005).
27. Xu, C. *et al.* Efficient generation and cryopreservation of cardiomyocytes derived from human embryonic stem cells. *Regen. Med.* **6**, 53–66 (2011).
28. Shiotani, M. *et al.* Practical application of guinea pig telemetry system for QT evaluation. *J. Toxicol. Sci.* **30**, 239–247 (2005).
29. Walker, M. J. *et al.* The Lambeth Conventions: guidelines for the study of arrhythmias in ischaemia infarction, and reperfusion. *Cardiovasc. Res.* **22**, 447–455 (1988).
30. Gutstein, D. E., Danik, S. B., Sereysky, J. B., Morley, G. E. & Fishman, G. I. Subdiaphragmatic murine electrophysiological studies: sequential determination of ventricular refractoriness and arrhythmia induction. *Am. J. Physiol. Heart Circ. Physiol.* **285**, H1091–H1096 (2003).
31. Biermann, M. *et al.* Differential effects of cytochalasin D and 2,3 butanedione monoxime on isometric twitch force and transmembrane action potential in isolated ventricular muscle: implications for optical measurements of cardiac repolarization. *J. Cardiovasc. Electrophysiol.* **9**, 1348–1357 (1998).
32. Laurita, K. R. & Singal, A. Mapping action potentials and calcium transients simultaneously from the intact heart. *Am. J. Physiol. Heart Circ. Physiol.* **280**, H2053–H2060 (2001).
33. Fedorov, V. V. *et al.* Application of blebbistatin as an excitation-contraction uncoupler for electrophysiologic study of rat and rabbit hearts. *Heart Rhythm* **4**, 619–626 (2007).
34. Kolega, J. Phototoxicity and photoinactivation of blebbistatin in UV and visible light. *Biochem. Biophys. Res. Commun.* **320**, 1020–1025 (2004).
35. Asfour, H., Swift, L. M., Sarvazyan, N., Doroslovacki, M. & Kay, M. W. Signal decomposition of transmembrane voltage-sensitive dye fluorescence using a multiresolution wavelet analysis. *IEEE Trans. Biomed. Eng.* **58**, 2083–2093 (2011).
36. Kay, M., Swift, L., Martell, B., Arutunyan, A. & Sarvazyan, N. Locations of ectopic beats coincide with spatial gradients of NADH in a regional model of low-flow reperfusion. *Am. J. Physiol. Heart Circ. Physiol.* **294**, H2400–H2405 (2008).
37. Swift, L. *et al.* Controlled regional hypoperfusion in Langendorff heart preparations. *Physiol. Meas.* **29**, 269–279 (2008).
38. Fernandes, S. *et al.* Human embryonic stem cell-derived cardiomyocytes engraft but do not alter cardiac remodeling after chronic infarction in rats. *J. Mol. Cell. Cardiol.* **49**, 941–949 (2010).
39. Laflamme, M. A. *et al.* Formation of human myocardium in the rat heart from human embryonic stem cells. *Am. J. Pathol.* **167**, 663–671 (2005).
40. Kuznetsov, S. A. *et al.* Circulating skeletal stem cells. *J. Cell Biol.* **153**, 1133–1140 (2001).

Averting biodiversity collapse in tropical forest protected areas

A list of the authors and their affiliations appears at the end of the paper.

The rapid disruption of tropical forests probably imperils global biodiversity more than any other contemporary phenomenon^{1–3}. With deforestation advancing quickly, protected areas are increasingly becoming final refuges for threatened species and natural ecosystem processes. However, many protected areas in the tropics are themselves vulnerable to human encroachment and other environmental stresses^{4–9}. As pressures mount, it is vital to know whether existing reserves can sustain their biodiversity. A critical constraint in addressing this question has been that data describing a broad array of biodiversity groups have been unavailable for a sufficiently large and representative sample of reserves. Here we present a uniquely comprehensive data set on changes over the past 20 to 30 years in 31 functional groups of species and 21 potential drivers of environmental change, for 60 protected areas stratified across the world's major tropical regions. Our analysis reveals great variation in reserve 'health': about half of all reserves have been effective or performed passably, but the rest are experiencing an erosion of biodiversity that is often alarmingly widespread taxonomically and functionally. Habitat disruption, hunting and forest-product exploitation were the strongest predictors of declining reserve health. Crucially, environmental changes immediately outside reserves seemed nearly as important as those inside in determining their ecological fate, with changes inside reserves strongly mirroring those occurring around them. These findings suggest that tropical protected areas are often intimately linked ecologically to their surrounding habitats, and that a failure to stem broad-scale loss and degradation of such habitats could sharply increase the likelihood of serious biodiversity declines.

Tropical forests are the biologically richest ecosystems on Earth^{1–3}. Growing concerns about the impacts of anthropogenic pressures on tropical biodiversity and natural ecosystem services have led to increases in the number and extent of protected areas across the tropics¹⁰. However, much remains unknown about the likelihood of biodiversity persisting in such protected areas. Remote-sensing technologies offer a bird's-eye view of tropical forests and provide many important insights^{6,11–13}, but are largely unable to discern crucial on-the-ground changes in forest biodiversity and ecological functioning¹⁴.

To appraise both the ecological integrity and threats for tropical protected areas on a global scale, we conducted a systematic and uniquely comprehensive assessment of long-term changes within 60 protected areas stratified across the world's major tropical forest regions (Supplementary Fig. 1). To our knowledge, no other existing data set includes such a wide range of biodiversity and threat indicators for such a large and representative network of tropical reserves. Our study was motivated by three broad issues: whether tropical reserves will function as 'arks' for biodiversity and natural ecosystem processes; whether observed changes are mainly concordant or idiosyncratic among different protected areas; and what the principal predictors of reserve success or failure are, in terms of their intrinsic characteristics and drivers of change.

To conduct our study we amassed expert knowledge from 262 detailed interviews, focusing on veteran field biologists and environmental scientists who averaged nearly 2 decades of experience

(mean \pm s.d., 19.1 \pm 9.6 years) at each protected area. Each interviewed researcher completed a detailed 10-page questionnaire, augmented by a telephone or face-to-face interview (see Supplementary Information). The questionnaires focused on longer-term (approximately 20–30-year) changes in the abundance of 31 animal and plant guilds (trophically or functionally similar groups of organisms), which collectively have diverse and fundamental roles in forest ecosystems (Table 1). We also recorded data on 21 potential drivers of environmental change both inside each reserve and within a 3-km-wide buffer zone immediately surrounding it (Table 1).

Our sample of protected areas spans 36 nations and represents a geographically stratified and broadly representative selection of sites across the African, American and Asia-Pacific tropics (Supplementary Fig. 1). The reserves ranged from 160 ha to 3.6 million ha in size, but most (85%) exceeded 10,000 ha in area (median = 99,350 ha; lower decile = 7,000 ha; upper decile = 750,000 ha). The protected areas fall under various International Union for Conservation of Nature (IUCN) reserve classifications. Using data from the World Database on Protected Areas (<http://www.wdpa.org>), we found no significant difference ($P = 0.13$) in the relative frequency of high-protection (IUCN Categories I–IV), multiple-use (Categories V–VI) and

Table 1 | The 31 animal and plant guilds, and the 21 environmental drivers assessed both inside and immediately outside each protected area.

Guilds	Potential environmental drivers
Broadly forest-dependent guilds	
Apex predators	Changes in natural-forest cover
Large non-predatory species	Selective logging
Primates	Fires
Opportunistic omnivorous mammals	Hunting
Rodents	Harvests of non-timber forest products
Bats	Illegal mining
Understory insectivorous birds	Roads
Raptorial birds	Automobile traffic
Larger frugivorous birds	Exotic plantations
Larger game birds	Human population density
Lizards and larger reptiles	Livestock grazing
Venomous snakes	Air pollution
Non-venomous snakes	Water pollution
Terrestrial amphibians	Stream sedimentation
Stream-dwelling amphibians	Soil erosion
Freshwater fish	River & stream flows
Dung beetles	Ambient temperature
Army or driver ants	Annual rainfall
Aquatic invertebrates	Drought severity or intensity
Large-seeded old-growth trees	Flooding
Epiphytes	Windstorms
Other functional groups	
Ecological specialists	
Species requiring tree cavities	
Migratory species	
Disturbance-favouring guilds	
Lianas and vines	
Pioneer and generalist trees	
Exotic animal species	
Exotic plant species	
Disease-vectoring invertebrates	
Light-loving butterflies	
Human diseases	

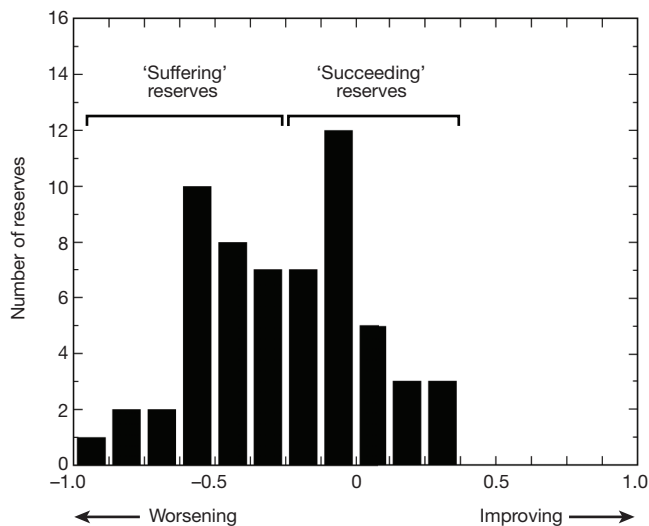


Figure 1 | Distribution of the 'reserve-health index' for 60 protected areas spanning the world's major tropical forest regions. This relative index averages changes in 10 well-studied guilds of animals and plants, including disturbance-avoiding and disturbance-favouring groups, over the past 20 to 30 years.

unclassified reserves between our sample of 60 reserves and all 16,038 reserves found in the same tropical nations (Supplementary Fig. 2). We also found no significant difference ($P = 0.08$) in the geographical isolation of our reserves (travel time to the nearest city with greater than 50,000 residents) relative to a random sample of 60 protected areas stratified across the same 36 nations (Supplementary Fig. 3).

We critically assessed the validity of our interview data by comparing them to 59 independent time-series data sets in which change in a single guild or environmental driver was assessed for one of our protected areas. Collectively, our meta-analysis included some data on 15 of the guilds, 13 of the drivers and 27 of the protected areas in our study (Supplementary Table 1). Most (86.4%) of the independent data sets supported our interview results, and in no case did an independent test report a trend opposite in sign to our interview-based findings.

Our analyses suggest that the most sensitive guilds in tropical protected areas include apex predators, large non-predatory vertebrates, bats, stream-dwelling amphibians, terrestrial amphibians, lizards and larger reptiles, non-venomous snakes, freshwater fish, large-seeded old-growth trees, epiphytes and ecological specialists (all $P < 0.0056$, with effect sizes ranging from -0.36 to -1.05 ; Supplementary Table 2). Several other groups were somewhat less vulnerable, including primates, understory insectivorous birds, large frugivorous birds,

raptorial birds, venomous snakes, species that require tree cavities, and migratory species (all $P < 0.05$, with effect sizes from -0.27 to -0.53). In addition, five groups increased markedly in abundance in the reserves, including pioneer and generalist trees, lianas and vines, invasive animals, invasive plants and human diseases (all $P < 0.0056$, with effect sizes from 0.44 to 1.17).

To integrate these disparate data, we generated a 'reserve-health index' that focused on 10 of the best-studied guilds (data for each available at $\geq 80\%$ of reserves), all of which seem to be sensitive to environmental changes in protected areas. Six of these are generally 'disturbance avoiders' (apex predators, large non-predatory vertebrates, primates, understory insectivorous birds, large frugivorous birds and large-seeded old-growth trees) and the remainder seem to be 'disturbance-favouring' groups (pioneer and generalist trees, lianas and vines, exotic animals and exotic plants). For each protected area, we averaged the mean values for each group, using negative values to indicate increases in abundance of the disturbance-favouring guilds.

The reserve-health index varied greatly among the different protected areas (Fig. 1). About four-fifths of the reserves had negative values, indicating some decline in reserve health. For 50% of all reserves this decline was relatively serious (mean score < -0.25), with the affected organisms being remarkable for their high functional and taxonomic diversity (Fig. 2). These included plants with varying growth forms and life-history strategies, and fauna that differed widely in body size, trophic level, foraging strategies, area needs, habitat use and other attributes. The remaining reserves generally exhibited much more positive outcomes for biodiversity (Fig. 2), although a few disturbance-favouring guilds, such as exotic plants and pioneer and generalist trees, often increased even within these areas.

An important predictor of reserve health was improving reserve management. According to our experts, reserves in which actual, on-the-ground protection efforts (see Supplementary Information) had increased over the past 20 to 30 years generally fared better than those in which protection had declined; a relationship that was consistent across all three of the world's major tropical regions (Fig. 3). Indeed, on-the-ground protection has increased in more than half of the reserves over the past 20 to 30 years, and this is assisting efforts to limit threats such as deforestation, logging, fires and hunting within these reserves (Supplementary Table 3), relative to areas immediately outside (Supplementary Table 4).

However, our findings show that protecting biodiversity involves more than just safeguarding the reserves themselves. In many instances, the landscapes and habitats surrounding reserves are under imminent threat^{5,6,15} (Fig. 4 and Supplementary Tables 3 and 4). For example, 85% of our reserves suffered declines in surrounding forest cover in the last 20 to 30 years, whereas only 2% gained surrounding forest. As shown by general linear models (Supplementary Table 5), such changes can seriously affect reserve biodiversity. Among the

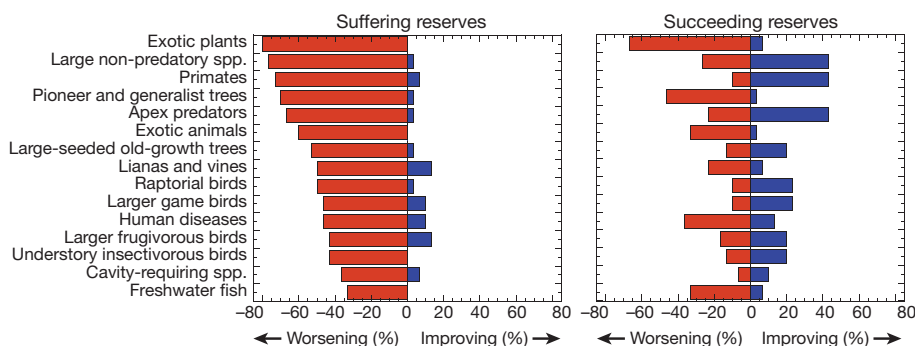


Figure 2 | Percentages of reserves that are worsening versus improving for key disturbance-sensitive guilds, contrasted between 'suffering' and 'succeeding' reserves (which are distinguished by having lower (< -0.25) versus higher (≥ -0.25) values for the reserve-health index, respectively). For disturbance-

favouring organisms such as exotic plants and animals, pioneer and generalist trees, lianas and vines, and human diseases, the reserve is considered to be worsening if the group increased in abundance. For any particular guild, reserves with missing or zero values (no trend) are not included.

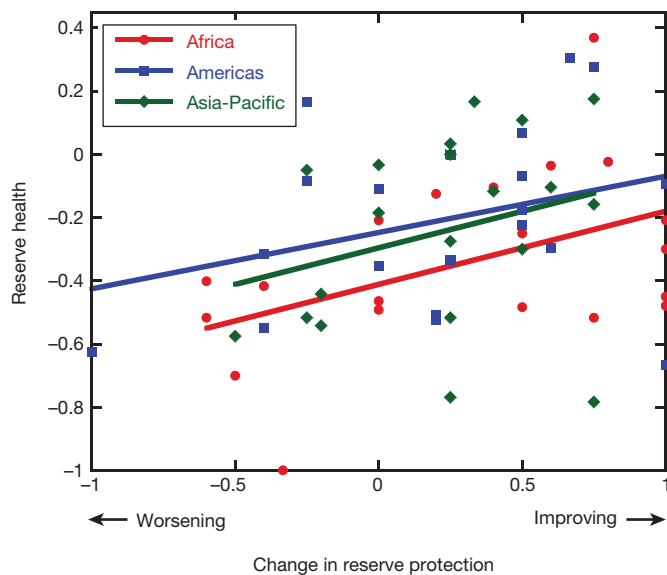


Figure 3 | Effects of improving on-the-ground protection on a relative index of reserve health. This positive relationship held across all three tropical continents (a general linear model showed that the protection term was the most effective predictor of reserve health (Akaike's information criterion weight, 0.595; deviance explained, 11.4%), with the addition of 'continent' providing only a small improvement in model fit (Akaike's information criterion weight, 0.317; deviance explained, 16.3%).

potential drivers of declining reserve health, three of the most important predictors involved ecological changes outside reserves (declining forest cover, increasing logging and increasing fires outside reserves; Supplementary Fig. 6). The remainder involved changes within reserves (particularly declining forest cover and increasing hunting, as well as increasing logging and harvests of non-timber forest products; Supplementary Table 5).

Thus, changes both inside and outside reserves determine their ecological viability, with forest disruption (deforestation, logging and fires), and overexploitation of wildlife and forest resources (hunting

and harvests of non-timber forest products) having the greatest direct negative impacts. Other environmental changes, such as air and water pollution, increases in human population densities and climatic change (changes in total rainfall, ambient temperature, droughts and windstorms) generally had weaker or more indirect effects over the last 20 to 30 years (Supplementary Table 5).

Environmental degradation occurring around a protected area could affect biodiversity in many ways, such as by increasing reserve isolation, area and edge effects^{15–19}. However, we discovered that its effects are also more insidious: they strongly predispose the reserve itself to similar kinds of degradation. Nearly all (19 of 21) of the environmental drivers had positive slopes when comparing their direction and magnitude inside versus outside reserves (Fig. 5). Among these, 13 were significant even with stringent Bonferroni corrections ($P < 0.0071$) and 17 would have been significant if tested individually ($P < 0.05$). As expected, the associations were strongest for climate parameters but were also strong for variables describing air and water pollution, stream sedimentation, hunting, mining, harvests of non-timber forest products and fires. To a lesser extent, trends in forest cover, human populations, road expansion and automobile traffic inside reserves also mirror those occurring outside reserves (Fig. 5).

Our findings signal that the fates of tropical protected areas will be determined by environmental changes both within and around the reserves, and that pressures inside reserves often closely reflect those occurring around them. For many reasons, larger reserves should be more resilient to such changes^{15–22}, although we found that removing the effects of reserve area statistically did not consistently weaken the correlations between changes inside versus outside protected areas (Supplementary Table 6).

Our study reveals marked variability in the health of tropical protected areas. It indicates that the best strategy for maintaining biodiversity within tropical reserves is to protect them against their major proximate threats, particularly habitat disruption and overharvesting. However, it is not enough to confine such efforts to reserve interiors while ignoring their surrounding landscapes, which are often being rapidly deforested, degraded and overhunted^{5,6,13,15} (Fig. 5). A failure to limit interrelated internal and external threats could predispose reserves to ecological decay, including a taxonomically and functionally

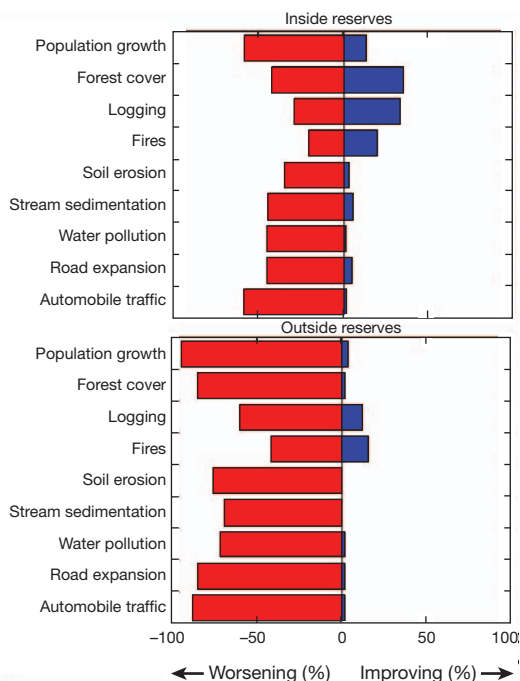
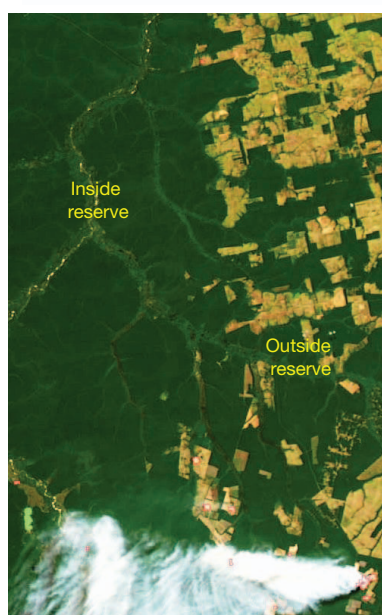


Figure 4 | Comparison of ecological changes inside versus outside protected areas, for selected environmental drivers. The image is an example of the strong distinction in disturbance inside versus outside a reserve. The bars show the percentages of reserves with improving versus worsening conditions.

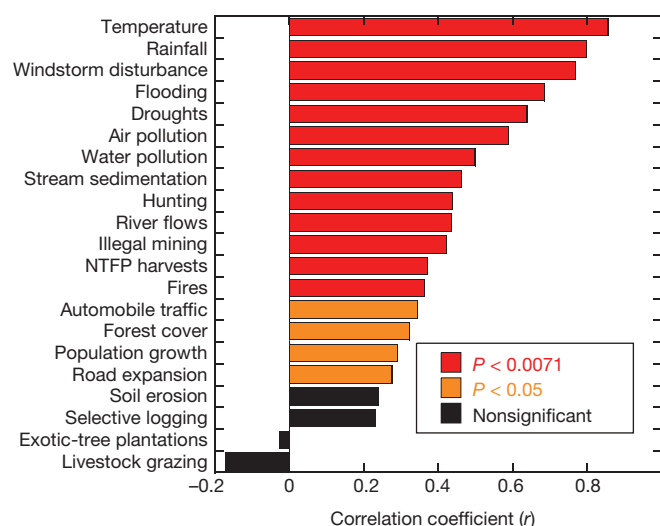


Figure 5 | Pearson correlations comparing the direction and strength of 21 environmental drivers inside versus outside tropical protected areas. NTFP, non-timber forest products.

sweeping array of changes in species communities (Fig. 2) and an erosion of fundamental ecosystem processes^{16,18,23}.

Protected areas are a cornerstone of efforts to conserve tropical biodiversity^{3,4,13,21}. It is not our intent to diminish their crucial role but to highlight growing challenges that could threaten their success. The vital ecological functions of wildlife habitats surrounding protected areas create an imperative, wherever possible, to establish sizeable buffer zones around reserves, maintain substantial reserve connectivity to other forest areas and promote lower-impact land uses near reserves by engaging and benefiting local communities^{4,15,24–27}. A focus on managing both external and internal threats should also increase the resilience of biodiversity in reserves to potentially serious climatic change^{28–30} in the future.

METHODS SUMMARY

Our interview protocol, rationale, questionnaire and data analyses are detailed in the Supplementary Information. We selected protected areas broadly to span the African, American and Asia-Pacific tropics (Supplementary Fig. 1), focusing on sites with mostly tropical or subtropical forest that had at least 10 refereed publications and 4–5 researchers with long-term experience who could be identified and successfully interviewed.

We devised a robust and relatively simple statistical approach to assess temporal changes in the abundance of each guild and in each potential environmental driver across our reserve network (see Supplementary Information). In brief, this involved asking each expert whether each variable had markedly increased, remained stable or markedly declined for each reserve. These responses were scored as 1, 0 and –1, respectively. For each response, the expert was also asked to rank their degree of confidence in their knowledge. After discarding responses with lower confidence, scores from the individual experts at each site were pooled to generate a mean value (ranging from –1.0 to 1.0) to estimate the long-term trend for each variable.

The means for each variable across all 60 sites were then pooled into a single data distribution. We used bootstrapping (resampling with replacement; 100,000 iterations) to generate confidence intervals for the overall mean of the data distribution. If the confidence intervals did not overlap zero, then we interpreted the trend as being non-random. Because we tested many different guilds, we used a stringent Bonferroni correction ($P \leq 0.0056$) to reduce the likelihood of Type I statistical errors, although we also identified guilds that showed evidence of trends ($P \leq 0.05$) if tested individually. For comparison, we estimated effect sizes (bootstrapped mean divided by s.d., with negative values indicating declines) for changes in guild abundances and for potential drivers inside and outside reserves (Supplementary Tables 2–4).

Received 24 February; accepted 14 June 2012.

Published online 25 July; corrected online 12 September 2012 (see full-text HTML version for details).

1. Pimm, S. L. & Raven, P. R. Biodiversity: extinction by numbers. *Nature* **403**, 843–845 (2000).

2. Bradshaw, C. J. A., Sodhi, N. S. & Brook, B. W. Tropical turmoil—a biodiversity tragedy in progress. *Front. Ecol. Environ.* **7**, 79–87 (2009).
3. Gibson, L. *et al.* Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature* **478**, 378–381 (2011).
4. Bruner, A. G., Gullison, R., Rice, R. & da Fonseca, G. Effectiveness of parks in protecting tropical biodiversity. *Science* **291**, 125–128 (2001).
5. Curran, L. M. *et al.* Lowland forest loss in protected areas of Indonesian Borneo. *Science* **303**, 1000–1003 (2004).
6. DeFries, R., Hansen, A., Newton, A. C. & Hansen, M. C. Increasing isolation of protected areas in tropical forests over the past twenty years. *Ecol. Appl.* **15**, 19–26 (2005).
7. Lovejoy, T. E. Protected areas: A prism for a changing world. *Trends Ecol. Evol.* **21**, 329–333 (2006).
8. Possingham, H. P., Wilson, K. A., Andelman, S. J. & Vynne, C. H. In *Principles of Conservation Biology* (eds Groom, M. J., Meffe, G. K. & Carroll, C. R.) (Sinauer, 2006).
9. Joppa, L. N., Loarie, S. & Pimm, S. L. On the protection of “protected areas”. *Proc. Natl Acad. Sci. USA* **105**, 6673–6678 (2008).
10. Jenkins, C. N. & Joppa, L. Expansion of the global terrestrial protected area system. *Biol. Conserv.* **142**, 2166–2174 (2009).
11. Asner, G. P. *et al.* Selective logging in the Brazilian Amazon. *Science* **310**, 480–482 (2005).
12. Wright, S. J., Sanchez-Azofeifa, G., Portillo-Quintero, C. & Davies, D. Poverty and corruption compromise tropical forest reserves. *Ecol. Appl.* **17**, 1259–1266 (2007).
13. Adeney, J. M., Christensen, N. & Pimm, S. L. Reserves protect against deforestation fires in the Amazon. *PLoS ONE* **4**, e5014 (2009).
14. Peres, C. A., Barlow, J. & Laurance, W. F. Detecting anthropogenic disturbance in tropical forests. *Trends Ecol. Evol.* **21**, 227–229 (2006).
15. Hansen, A. J. & DeFries, R. Ecological mechanisms linking protected areas to surrounding lands. *Ecol. Appl.* **17**, 974–988 (2007).
16. Laurance, W. F. *et al.* Biomass collapse in Amazonian forest fragments. *Science* **278**, 1117–1118 (1997).
17. Woodroffe, R. & Ginsberg, J. R. Edge effects and the extinction of populations inside protected areas. *Science* **280**, 2126–2128 (1998).
18. Terborgh, J. *et al.* Ecological meltdown in predator-free forest fragments. *Science* **294**, 1923–1926 (2001).
19. Laurance, W. F. *et al.* The fate of Amazonian forest fragments: a 32-year investigation. *Biol. Conserv.* **144**, 56–67 (2011).
20. Brooks, T. M., Pimm, S. L. & Oyugi, J. O. Time lag between deforestation and bird extinction in tropical forest fragments. *Conserv. Biol.* **13**, 1140–1150 (1999).
21. Peres, C. A. Why we need megareserves in Amazonia. *Conserv. Biol.* **19**, 728–733 (2005).
22. Maiorano, L., Falcucci, A. & Boitani, L. Size-dependent resistance of protected areas to land-use change. *Proc. R. Soc. B* **275**, 1297–1304 (2008).
23. Estes, J. A. *et al.* Trophic downgrading of Planet Earth. *Science* **333**, 301–306 (2011).
24. Wells, M. P. & McShane, T. O. Integrating protected area management with local needs and aspirations. *Ambio* **33**, 513–519 (2004).
25. Scherl, L. M. *et al.* Can Protected Areas Contribute to Poverty Reduction? Opportunities and Limitations (IUCN, 2004).
26. Chan, K. M. A. & Daily, G. C. The payoff of conservation investments in tropical countries. *Proc. Natl Acad. Sci. USA* **105**, 19342–19347 (2008).
27. Porter-Bolland, L. *et al.* Community-managed forests and protected areas: an assessment of their conservation effectiveness across the tropics. *For. Ecol. Manage.* **256**, 6–17 (2012).
28. Thomas, C. D. *et al.* Extinction risk from climate change. *Nature* **427**, 145–148 (2004).
29. Sekercioglu, C. H., Schneider, S. H., Fay, J. P. & Loarie, S. R. Climate change, elevational range shifts, and bird extinctions. *Conserv. Biol.* **22**, 140–150 (2008).
30. Shoo, L. P. *et al.* Targeted protection and restoration to conserve tropical biodiversity in a warming world. *Glob. Change Biol.* **17**, 186–193 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The study was supported by James Cook University, the Smithsonian Tropical Research Institute, an Australian Laureate Fellowship (to W.F.L.) and NSF grant RCN-0741956. We thank A. Bruner, R. A. Butler, G. R. Clements, R. Condit, C. N. Cook, S. Goosem, J. Geldmann, L. Joppa, S. L. Pimm and O. Venter for comments.

Author Contributions W.F.L. conceived the study and coordinated its design, analysis and manuscript preparation. D.C.U., J.R. and M.K. conducted the interviews; C.J.A.B. assisted with data analysis and some writing; and S.P.S., S.G.L., M.C. and W.L. organized data or collected metadata. The remaining authors provided detailed interviews on protected areas and offered feedback on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.F.L. (bill.laurance@jcu.edu.au).

William F. Laurance^{1,2}, D. Carolina Useche², Julio Rendeiro², Margareta Kalka², Corey J. A. Bradshaw³, Sean P. Sloan¹, Susan G. Laurance¹, Mason Campbell¹, Kate

- Abernethy⁴, Patricia Alvarez⁵, Victor Arroyo-Rodriguez⁶, Peter Ashton⁷, Julieta Benítez-Malvido⁸, Allard Blom⁹, Kadiri S. Bobo⁹, Charles H. Cannon¹⁰, Min Cao¹⁰, Richard Carroll⁸, Colin Chapman¹¹, Rosamond Coates¹², Marina Cords¹³, Finn Danielsen¹⁴, Bart De Dijn¹⁵, Eric Dinerstein⁸, Maureen A. Donnelly¹⁶, David Edwards¹, Felicity Edwards¹, Nina Farwig¹⁷, Peter Fashing¹⁸, Pierre-Michel Forget¹⁹, Mercedes Foster²⁰, George Gale²¹, David Harris²², Rhett Harrison¹⁰, John Hart²³, Sarah Karpanty²⁴, W. John Kress²⁵, Jagdish Krishnaswamy²⁶, Willis Logsdon¹, Jon Lovett²⁷, William Magnusson²⁸, Fiona Maisels^{4,29}, Andrew R. Marshall³⁰, Deedra McClearn³¹, Divya Mudappa³², Martin R. Nielsen³³, Richard Pearson³⁴, Nigel Pitman⁵, Jan van der Ploeg³⁵, Andrew Plumpton³⁶, John Poulsen³⁷, Mauricio Quesada⁶, Hugo Rainey²⁹, Douglas Robinson³⁸, Christiane Roetgers¹, Francesco Rovero³⁹, Frederick Scatena⁴⁰, Christian Schulze⁴¹, Douglas Sheil⁴², Thomas Struhsaker⁵, John Terborgh⁵, Duncan Thomas³⁸, Robert Timm⁴³, J. Nicolas Urbina-Cardona⁴⁴, Karthikeyan Vasudevan⁴⁵, S. Joseph Wright⁴⁶, Juan Carlos Arias-G.⁴⁶, Luzmila Arroyo⁴⁷, Mark Ashton⁴⁸, Philippe Auzel¹¹, Dennis Babaasa⁴⁹, Fred Babweteera⁵⁰, Patrick Baker⁵¹, Olaf Banki⁵², Margot Bass⁵³, Inogwabini Billa-Issa⁵⁴, Stephen Blake²⁹, Warren Brockelman⁵⁵, Nicholas Brokaw⁵⁶, Carsten A. Brühl⁵⁷, Sarayudh Bunyavejchewin⁵⁸, Jung-Tai Chao⁵⁹, Jerome Chave⁶⁰, Ravi Chellam⁶¹, Connie J. Clark⁵, José Clavijo⁶², Robert Congdon³⁴, Richard Corlett⁶³, H. S. Dattaraja⁶⁴, Chittaranjan Dave⁶⁵, Glyn Davies⁶⁶, Beatriz de Mello Beisiegel⁶⁷, Rosa de Nazaré Paes da Silva⁶⁸, Anthony Di Fiore⁶⁹, Arvin Diesmos⁷⁰, Rodolfo Dirzo⁷¹, Diane Doran-Sheehy⁷², Mitchell Eaton⁷³, Louise Emmons²⁵, Alejandro Estrada¹², Corneille Ewango⁷⁴, Linda Fedigan⁷⁵, François Feer¹⁹, Barbara Fruth⁷⁶, Jacalyn Giacalone Willis⁷⁷, Uromi Goodale⁷⁸, Steven Goodman⁷⁹, Juan C. Guix⁸⁰, Paul Guthiga⁸¹, William Haber⁸², Keith Hamer⁸³, Ilka Herbringer⁸⁴, Jane Hill³⁰, Zhongliang Huang⁸⁵, I Fang Sun⁸⁶, Kalan Ickes⁸⁷, Akira Itoh⁸⁸, Natália Ivauskas⁸⁹, Betsy Jackes³⁴, John Janovec⁹⁰, Daniel Janzen⁴⁰, Mo Jiangming⁹¹, Chen Jin¹⁰, Trevor Jones⁹², Hermes Justiniano⁹³, Elisabeth Kalko⁹⁴, Aventino Kasangaki⁹⁵, Timothy Killeen⁹⁶, Hen-biao King⁹⁷, Erik Klop⁹⁸, Cheryl Knott⁹⁹, Inza Koné¹⁰⁰, Enoke Kudavidanage¹⁰¹, José Lahoz da Silva Ribeiro¹⁰¹, John Lattke¹⁰², Richard Laval¹⁰³, Robert Lawton¹⁰⁴, Miguel Leal¹⁰⁵, Mark Leighton¹⁰⁶, Miguel Lentino¹⁰⁷, Cristiane Leone¹⁰⁸, Jeremy Lindsell¹⁰⁹, Lee Ling-Ling¹¹⁰, K. Eduard Linsenmair¹¹¹, Elizabeth Losos¹¹², Ariel Lugo¹¹³, Jeremiah Lwanga¹¹⁴, Andrew L. Mack¹¹⁵, Marclia Martins¹¹⁶, W. Scott McGraw¹¹⁷, Roan McNab¹¹⁸, Luciano Montag¹¹⁹, Jo Myers Thompson¹²⁰, Jacob Nabe-Nielsen¹²¹, Michiko Nakagawa¹²², Sanjay Nepal¹²³, Marilyn Norconk¹²⁴, Vojtech Novotny¹²⁵, Sean O'Donnell¹²⁶, Muse Opiang¹²⁷, Paul Ouboter¹²⁸, Kenneth Parker¹²⁹, N. Parthasarathy¹³⁰, Kátia Pisciotto¹³¹, Dewi Prawiradilaga¹³², Catherine Pringle¹³³, Subaraj Rajathurai¹³⁴, Ulrich Reichard¹³⁵, Gay Reinartz¹³⁶, Katherine Renton¹³⁷, Glen Reynolds¹³⁸, Vernon Reynolds¹³⁹, Erin Riley¹⁴⁰, Mark-Oliver Rödel¹⁴¹, Jessica Rothman¹⁴², Philip Round¹⁴³, Shoko Sakai¹⁴⁴, Tania Sanaiotti²⁸, Tommaso Savini²¹, Gertrud Schaab¹⁴⁵, John Seidensticker¹⁴⁶, Alhaji Siaka¹⁴⁷, Miles R. Silman¹⁴⁸, Thomas B. Smith¹⁴⁹, Samuel Soares de Almeida¹⁵⁰, Navjot Sodhi⁶³, Craig Stanford¹⁵¹, Kristine Stewart¹⁵², Emma Stokes²⁹, Kathryn E. Stoner¹⁵³, Raman Sukumar¹⁵⁴, Martin Surbeck⁷⁶, Mathias Tobler⁹⁰, Teja Tscharnkte¹⁵⁵, Andrea Turkalo¹⁵⁶, Govindaswamy Umaphathy¹⁵⁷, Merlijn van Weerd³⁵, Jorge Vega Rivera¹³⁷, Meena Venkataraman¹⁵⁸, Linda Venn¹⁵⁹, Carlos Vereza¹⁶⁰, Carolina Volkmer de Castilho¹⁶¹, Matthias Walter¹⁵⁵, Benjamin Wang¹⁴⁹, David Watts⁴⁸, William Weber²⁹, Paige West¹³, David Whitacre¹⁶², Ken Whitney¹⁶³, David Wilkie²⁹, Stephen Williams³⁴, Debra D. Wright¹¹⁵, Patricia Wright¹⁶⁴, Lu Xiankai⁹¹, Pralad Yonzon¹⁶⁵ & Franky Zamzani¹⁶⁶
- ¹Centre for Tropical Environmental and Sustainability Science (TESS) and School of Marine and Tropical Biology, James Cook University, Cairns, Queensland 4878, Australia. ²Smithsonian Tropical Research Institute, Balboa, Ancón, Panama. ³School of Earth and Environmental Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia. ⁴Stirling University, Stirling FK9 4LA, UK. ⁵Duke University, Durham, North Carolina 27705, USA. ⁶Universidad Nacional Autónoma de México (UNAM), Morelia, Mexico. ⁷Royal Botanic Gardens, Kew, Richmond TW9 3AB, UK. ⁸World Wildlife Fund (WWF), Washington DC 20037, USA. ⁹University of Dschang, Dschang, Cameroon. ¹⁰Xishuangbanna Tropical Botanical Garden, Yunnan 666303, People's Republic of China. ¹¹McGill University, Montreal H3A 2T7, Canada. ¹²Estación de Biología Tropical Los Tuxtlas, Universidad Nacional Autónoma de México, Veracruz 95701, Mexico. ¹³Columbia University, New York, New York 10027, USA. ¹⁴Nordic Foundation for Development and Ecology, DK-1159 Copenhagen, Denmark. ¹⁵Bart De Dijn Environmental Consultancy, Paramaribo, Suriname. ¹⁶Florida International University, Miami, Florida 33199, USA. ¹⁷Philipps-Universität Marburg, Marburg 35043, Germany. ¹⁸California State University, Fullerton, California 92834, USA. ¹⁹Museum National d'Histoire Naturelle, 91800 Brunoy, France. ²⁰US Geological Survey, Smithsonian Institution, Washington DC 20013, USA. ²¹King Mongkut's University of Technology Thonburi, Bangkok 10150, Thailand. ²²Royal Botanic Garden, Edinburgh, Scotland EH3 5LR, UK. ²³Tshuapa-Lomami-Lualaba Project, Kinshasa, Democratic Republic of Congo. ²⁴Virginia Tech University, Blacksburg, Virginia 24061, USA. ²⁵National Museum of Natural History, Smithsonian Institution, Washington DC 20013, USA. ²⁶Ashoka Trust for Research in Ecology and the Environment (ATREE), Bangalore 560064, India. ²⁷University of Twente, Enschede, Netherlands. ²⁸Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Amazonas 69011-970, Brazil. ²⁹Wildlife Conservation Society, Bronx, New York 10460, USA. ³⁰University of York, Heslington, York YO10 5DD, UK. ³¹La Selva Biological Station, San Pedro, Costa Rica. ³²Nature Conservation Foundation, Mysore 570 002, India. ³³University of Copenhagen, Copenhagen, Denmark. ³⁴James Cook University, Townsville, Queensland 4811, Australia. ³⁵Leiden University, Leiden, Netherlands. ³⁶Wildlife Conservation Society, Kampala, Uganda. ³⁷Woods Hole Research Center, Falmouth, Massachusetts 02540, USA. ³⁸Oregon State University, Corvallis, Oregon 97331, USA. ³⁹Museo delle Scienze, 38122 Trento, Italy. ⁴⁰University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴¹University of Vienna, 1030 Vienna, Austria. ⁴²Bwindi Impenetrable National Park, Kabale, Uganda. ⁴³University of Kansas, Lawrence, Kansas 66045, USA. ⁴⁴Pontificia Universidad Javeriana, Bogotá, Colombia. ⁴⁵Wildlife Institute of India, Dehradun, India. ⁴⁶Unidad de Parques Nacionales Naturales de Colombia, Bogotá, Colombia. ⁴⁷Museo de Historia Natural Noel Kempff, Santa Cruz, Bolivia. ⁴⁸Yale University, New Haven, Connecticut 06511, USA. ⁴⁹Institute of Tropical Forest Conservation, Kabale, Uganda. ⁵⁰Budongo Conservation Field Station, Masindi, Uganda. ⁵¹Monash University, Melbourne, Victoria 3800, Australia. ⁵²Utrecht University, Utrecht, Netherlands. ⁵³Finding Species, Takoma Park, Maryland 20912, USA. ⁵⁴University of Kent, Kent CT2 7NZ, UK. ⁵⁵Mahidol University Salaya, Nakhon Pathom 73170, Thailand. ⁵⁶University of Puerto Rico, San Juan 00936, Puerto Rico. ⁵⁷University Koblenz-Landau, D-76829 Landau, Germany. ⁵⁸Department of National Parks, Chatuchak, Bangkok 10900, Thailand. ⁵⁹Taiwan Forestry Research Institute, Taipei 10066, Taiwan. ⁶⁰Université Paul Sabatier, Toulouse, France. ⁶¹Wildlife Conservation Society, Bangalore 560070, India. ⁶²Universidad Central de Venezuela, Aragua, Venezuela. ⁶³National University of Singapore, Singapore 117543. ⁶⁴Indian Institute of Science, Bangalore 560012, India. ⁶⁵World Wide Fund for Nature (WWF), New Delhi 110003, India. ⁶⁶World Wide Fund for Nature (WWF), Surrey GU7 1XR, UK. ⁶⁷Instituto Chico Mendes de Conservação de Biodiversidade, Atibaia, São Paulo 12952-011, Brazil. ⁶⁸O Conselho Regional de Engenharia, Arquitetura e Agronomia do Pará, Belém, Pará, Brazil. ⁶⁹University of Texas, Austin, Texas 78712, USA. ⁷⁰National Museum of the Philippines, Manila, Philippines. ⁷¹Stanford University, Stanford, California 94305, USA. ⁷²State University of New York at Stony Brook, Stony Brook, New York 11794, USA. ⁷³University of Colorado, Boulder, Colorado 80309, USA. ⁷⁴Wildlife Conservation Society, Kinshasa, Democratic Republic of Congo. ⁷⁵University of Calgary, Alberta T2N 1N4, Canada. ⁷⁶Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ⁷⁷Montclair State University, Montclair, New Jersey 07043, USA. ⁷⁸University of California, San Diego, California 92093, USA. ⁷⁹Field Museum of Natural History, Chicago, Illinois 60605, USA. ⁸⁰Universitat de Barcelona, 08028 Barcelona, Spain. ⁸¹Kenya Institute for Public Policy Research and Analysis, Nairobi, Kenya. ⁸²Missouri Botanical Garden, St. Louis, Missouri 63166, USA. ⁸³University of Leeds, Leeds LS2 9JT, UK. ⁸⁴Wild Chimpanzee Foundation, Abidjan 23, Côte d'Ivoire. ⁸⁵Dinghushan Biosphere Reserve, Zhaoqing, People's Republic of China. ⁸⁶Tungshai University, Taichung 407, Taiwan. ⁸⁷Clemson University, Clemson, South Carolina 29634, USA. ⁸⁸Osaka City University, Osaka 558-8585, Japan. ⁸⁹Instituto Florestal, São Paulo, São Paulo 02377-000, Brazil. ⁹⁰Botanical Research Institute of Texas, Fort Worth, Texas 76107, USA. ⁹¹South China Botanical Garden, Guangzhou 510650, People's Republic of China. ⁹²Anglia Ruskin University, Cambridge CB1 1PT, UK. ⁹³Fundación para la Conservación del Bosque Chiquitano, Bolivia. ⁹⁴University of Ulm, 89069 Ulm, Germany. ⁹⁵Mbarara University of Science and Technology, Mbarara, Uganda. ⁹⁶Conservation International, Arlington, Virginia 22202, USA. ⁹⁷Society of Subtropical Ecology, Taipei, Taiwan. ⁹⁸Royal Haskoning, Water and Ecology Group, Groningen, Netherlands. ⁹⁹Boston University, Boston, Massachusetts 02215, USA. ¹⁰⁰Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire. ¹⁰¹Universidade Estadual de Londrina, Londrina, Paraná, Brazil. ¹⁰²Universidad Central de Venezuela, Caracas, Venezuela. ¹⁰³The Bat Jungle, Monteverde, Costa Rica. ¹⁰⁴University of Alabama, Huntsville, Alabama 35899, USA. ¹⁰⁵Boite Postale 7847, Libreville, Gabon. ¹⁰⁶US Warren Road, Framingham, Massachusetts 01702, USA. ¹⁰⁷Colección Ornitológica Phelps, Caracas, Venezuela. ¹⁰⁸Parque Estadual Horto Florestal, São Paulo, São Paulo 02377-000, Brazil. ¹⁰⁹Royal Society for the Protection of Birds, Sandy SG19 2DL, UK. ¹¹⁰National Taiwan University, Taipei, Taiwan. ¹¹¹University of Würzburg, Biocenter, D97074 Würzburg, Germany. ¹¹²Organization for Tropical Studies, Durham, North Carolina 27705, USA. ¹¹³USDA International Institute of Tropical Forestry, Río Piedras, Puerto Rico 00926. ¹¹⁴Makerere University, Kampala, Uganda. ¹¹⁵Green Capacity Inc., New Florence, Pennsylvania 15944, USA. ¹¹⁶Museu Paraense Emílio Goeldi, Belém, Pará 66040-170, Brazil. ¹¹⁷Ohio State University, Columbus, Ohio 43210, USA. ¹¹⁸Wildlife Conservation Society, Flores, Guatemala. ¹¹⁹Universidad Federal do Pará, Belém, Pará 66040-170, Brazil. ¹²⁰Lukuru Wildlife Research Foundation, Kinshasa, Democratic Republic of Congo. ¹²¹Aarhus University, 4000 Roskilde, Denmark. ¹²²Nagoya University, Nagoya, Japan. ¹²³University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. ¹²⁴Kent State University, Kent, Ohio 44242, USA. ¹²⁵Institute of Entomology, Ceske Budejovice, Czech Republic. ¹²⁶University of Washington, Seattle, Washington 98195, USA. ¹²⁷PNG Institute of Biological Research, Goroka, Papua New Guinea. ¹²⁸University of Suriname, Paramaribo, Suriname. ¹²⁹113-3885 Richet Rd, Prince George, British Columbia V2K 2J2, Canada. ¹³⁰Pondicherry University, Puducherry 605-014, India. ¹³¹Fundação Florestal, São Paulo, São Paulo 02377-000, Brazil. ¹³²Research Centre for Biology, Cibinong 16911, Indonesia. ¹³³University of Georgia, Athens, Georgia 30602, USA. ¹³⁴Strix Wildlife Consultancy, Singapore. ¹³⁵Southern Illinois University, Carbondale, Illinois 62901, USA. ¹³⁶Zoological Society of Milwaukee, Milwaukee, Wisconsin 53226, USA. ¹³⁷Estación de Biología Chamela, Universidad Nacional Autónoma de México, Jalisco 48980, Mexico. ¹³⁸Danum Valley Field Centre, Sabah, Malaysia. ¹³⁹Oxford University, Oxford BN26 5UX, UK. ¹⁴⁰San Diego State University, San Diego, California 92182, USA. ¹⁴¹Museum für Naturkunde, Berlin, Germany. ¹⁴²City University of New York, New York 10065, USA. ¹⁴³Mahidol University, Bangkok 10400, Thailand. ¹⁴⁴Research Institute for Humanity and Nature, Kyoto, Japan. ¹⁴⁵Karlsruhe University of Applied Sciences, Karlsruhe, Germany. ¹⁴⁶National Zoological Park, Washington DC 20013, USA. ¹⁴⁷Gola Forest Programme, Kenema, Sierra Leone. ¹⁴⁸Wake Forest University, Winston-Salem, North Carolina 27106, USA. ¹⁴⁹University of California, Los Angeles, California 90095, USA. ¹⁵⁰Av. Maalhães Barata 376, Belém, Pará 66040-170, Brazil. ¹⁵¹University of Southern California, Los Angeles, California 90089, USA. ¹⁵²Institute of Applied Ethnobotany, Pompano Beach, Florida 33069, USA. ¹⁵³Texas A & M University, Kingsville, Texas 78363, USA. ¹⁵⁴Indian Institute of Science, Bangalore, India. ¹⁵⁵Georg-August-Universität, Göttingen, Germany. ¹⁵⁶Wildlife Conservation Society, Bangui, Central African Republic. ¹⁵⁷Centre for Cellular and Molecular Biology, Hyderabad, India. ¹⁵⁸701, Vesta B, Lodha Paradise, Thane, India. ¹⁵⁹Paluma Environmental Education Centre, Paluma, Queensland 4816, Australia. ¹⁶⁰Universidad Central de Venezuela, Maracay, Venezuela. ¹⁶¹Embrapa Roraima, Boa Vista, Roraima, Brazil. ¹⁶²Treasure Valley Math and Science Center, Boise, Idaho 83714, USA. ¹⁶³Rice University, Houston, Texas 77005, USA. ¹⁶⁴Stony Brook University, Stony Brook, New York 11794, USA. ¹⁶⁵Resources Himalaya Foundation, Kathmandu, Nepal. ¹⁶⁶Gunung Palung National Park, West Kalimantan, Indonesia.
- ‡Deceased.

Activity in motor–sensory projections reveals distributed coding in somatosensation

Leopoldo Petreanu^{1†}, Diego A. Gutnisky¹, Daniel Huber^{1†}, Ning-long Xu¹, Dan H. O'Connor¹, Lin Tian^{1†}, Loren Looger¹ & Karel Svoboda¹

Cortical-feedback projections to primary sensory areas terminate most heavily in layer 1 (L1) of the neocortex^{1,2}, where they make synapses with tuft dendrites of pyramidal neurons. L1 input is thought to provide ‘contextual’ information³, but the signals transmitted by L1 feedback remain uncharacterized. In the rodent somatosensory system, the spatially diffuse⁴ feedback projection from vibrissal motor cortex (vM1) to vibrissal somatosensory cortex (vS1, also known as the barrel cortex) may allow whisker touch to be interpreted in the context of whisker position to compute object location^{5,6}. When mice palpate objects with their whiskers to localize object features^{7,8}, whisker touch excites vS1⁹ and later vM1 in a somatotopic manner^{10–13}. Here we use axonal calcium imaging to track activity in vM1→vS1 afferents in L1 of the barrel cortex while mice performed whisker-dependent object localization. Spatially intermingled individual axons represent whisker movements, touch and other behavioural features. In a subpopulation of axons, activity depends on object location and persists for seconds after touch. Neurons in the barrel cortex thus have information to integrate movements and touches of multiple

whiskers over time, key components of object identification and navigation by active touch.

We trained head-fixed mice to perform a whisker-based object-localization task under a two-photon microscope with one whisker row (C row) (Fig. 1a)¹². In each trial, a pole was moved into one of several locations within reach of the whiskers (sampling period, approximately 1 s; Supplementary Fig. 1a), arranged along the anterior–posterior axis on one side of the head (Fig. 1b). Mice moved their whiskers to determine whether the pole was either in one of several ‘go’ locations or in a ‘no-go’ location. Automated whisker tracking¹⁴ measured whisker movements and shape. The force acting on the follicles, which underlies object localization^{8,15,16}, can be estimated from whisker shape changes induced by touch. A change in curvature at point *p* of the whisker is proportional to the force applied by the pole on the whisker¹⁶: $F \propto \Delta\kappa_p y_p$, where y_p is the bending stiffness at *p*. We thus present forces acting on the whiskers as the change in curvature, $\Delta\kappa$ (Supplementary Fig. 1b). Mice report their decision about object location with licking. Imaging was performed in trained mice (discriminability (d') average,

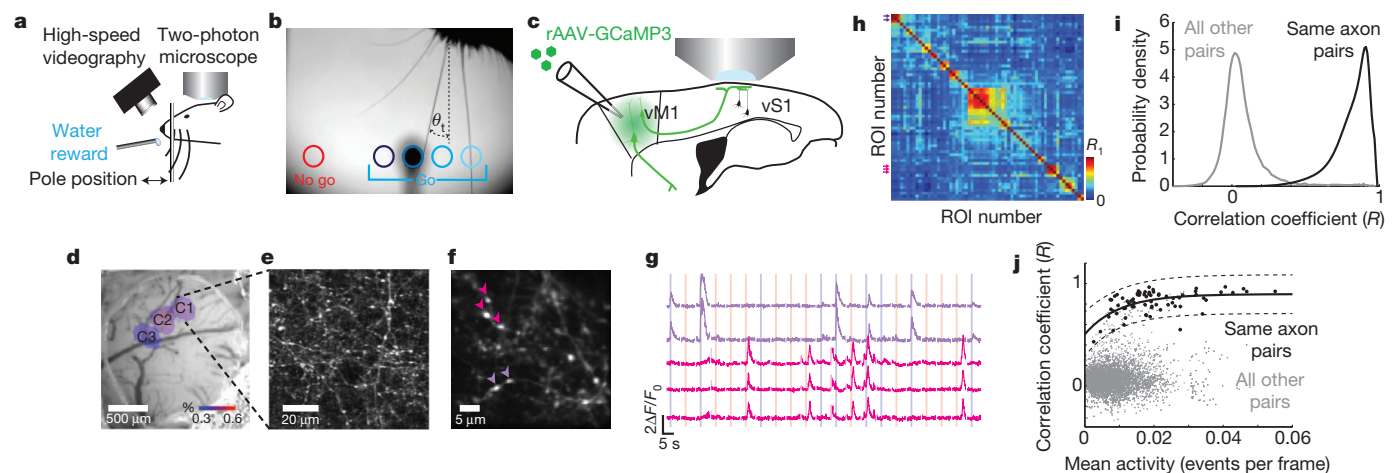


Figure 1 | Imaging activity in vM1→vS1 axons during whisker-based object localization. **a**, Head-fixed mouse performing an object-localization task under the microscope. **b**, One frame of a high-speed video sequence and typical pole positions corresponding to multiple go locations (hues of blue) and a single no-go location (red). Whisker touch was possible for all object locations for the long posterior whiskers. The azimuthal angle θ describes whisker position (θ_1 denotes azimuthal angle at first touch). **c**, vM1 neurons were infected with GCaMP3-expressing virus and their axons were imaged in L1 of vS1. **d**, Intrinsic optical-imaging signals corresponding to deflections of whiskers C1, C2 and C3, overlaid on a brightfield image of the vasculature. **e**, Two-photon images showing vM1 axons in the C1 barrel column. **f**, Field of view used for *in vivo* imaging showing varicosities (arrowheads correspond to varicosities used for

fluorescence time series in **g**). **g**, Fluorescence dynamics over multiple behavioural trials. Coloured bars indicate the sampling period, when the pole is within reach (blue, go trials; red, no-go trials). **h**, Matrix of correlation coefficients for activity in different varicosities in one field of view. Arrowheads correspond to varicosities marked in **f**. **i**, Distribution of correlation coefficients for all fields of view. Black line denotes pairs of varicosities on the same axon (78 varicosities, 31 axons) and grey line denotes all other varicosity pairs (51,325 varicosities). Correlation coefficients were computed over the entire session. **j**, Correlation coefficients as a function of mean activity level (black denotes pairs of varicosities on the same axon and grey denotes all other pairs). The solid line is an exponential fit ($y = a(1 - e^{-bx}) + c$) to the black circles and the dashed lines represent the 95% confidence interval.

¹Janelia Farm Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA. [†]Present addresses: Department for Basic Neurosciences, University of Geneva, Geneva CH-1211, Switzerland (D.H.); Champalimaud Neuroscience Programme, Lisbon 1400-038, Portugal (L.P.); Department of Biochemistry and Molecular Medicine, School of Medicine, University of California, Davis California 95616, USA (L.T.).

1.8, corresponding to 80% correct trials; range, 1.2–2.6) (Supplementary Fig. 1c–e).

To image activity in vM1→vS1 axons we expressed the genetically encoded calcium indicator GCaMP3 (ref. 17) in vM1 neurons^{5,12} and implanted a chronic imaging window over vS1 (ref. 18) (Fig. 1c). Imaging was in barrel columns with principal whiskers C1, C2 or C3, based on intrinsic signal imaging¹⁹ (Fig. 1d). Fluorescent axons were abundant in L1 (Fig. 1e, f and Supplementary Fig. 2a, b). GCaMP3-positive axonal varicosities appeared as bright spots along the axon backbone²⁰. Two-photon time-lapse images (field of view, $40\ \mu\text{m} \times 40\ \mu\text{m}$) were acquired using raster scanning (16 Hz) over 100–200 trials per imaging location (trial duration, 7 s) (Fig. 1f, g and Supplementary Fig. 1c, d).

Action potentials reliably invade cortical axonal arborizations and cause calcium accumulations in varicosities²¹, which harbour synaptic terminals. We measured GCaMP3-mediated fluorescence transients, corresponding to trains of action potentials (≥ 5 ; Supplementary Fig. 2c–f), in regions of interest (ROIs) containing individual varicosities (range, 35–80 varicosities per field of view; 17 fields of view; 6 mice). Fluorescence transients were detected in multiple ROIs within each field of view (range, 6–20 active varicosities per field of view) (Figs 1f and 2a). As expected, the fluorescence signals from varicosities belonging to the same axon were highly correlated, whereas correlations computed across all other pairs were low (Fig. 1g). These correlations allowed us to identify varicosities on the same axon, even in cases where the local arborization could not be reconstructed morphologically (Fig. 1h–j). Below we report the activity of one varicosity per axon to represent the activity of the entire axon.

Only active axons were analysed (Methods). This ensured that almost all (>99%) fluorescence signals reported neural activity rather than movement (Supplementary Fig. 3).

We aligned fluorescence time series with recordings of behavioural features, including $\Delta\kappa$, whisker movements (whisking) and lick rate, and grouped trials by trial type (Fig. 2). Axonal activity was concentrated around the sampling period, and correlated with specific behavioural variables. Some axons were active coincident with whisking before and during the sampling period (Fig. 2b–d; compare ROI1 and whisking amplitude), whereas other axons were mainly active in trials with strong whisker touch (compare ROI2 and $\Delta\kappa$). Some axons were primarily active during licking or could not obviously be explained by a single behavioural feature¹². Correlations between activity and behaviour were apparent across trials and within trials. These data indicate that individual vM1→vS1 axons represent specific behavioural features, but the representations are diverse across the population of axons.

We used a generalized form of regression (Random Forests, Methods) to quantify how behavioural features are represented by vM1→vS1 axons¹². Both behaviour and activity varied within trials, across individual trials and across trial types (Fig. 2b, c). For example, in some trials whisker touch was strong, whereas in other trials it was weak or absent (touch in no-go trials was typically the weakest; Supplementary Fig. 1b, e). In some trials mice whisked gingerly and late during the sampling period (Fig. 2b, trial 144), whereas in other trials they whipped their whiskers against the pole as soon as it came within reach (Fig. 2b, trial 145). Our algorithms exploit this variability and the large number of trials in our data set (Supplementary Fig. 1d)

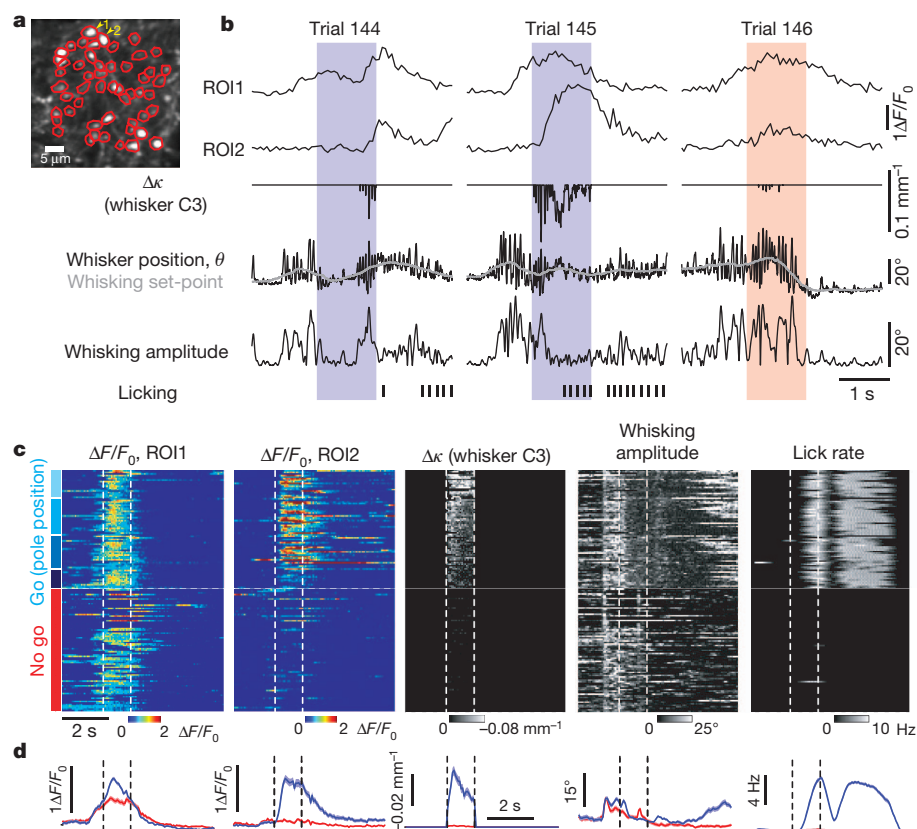


Figure 2 | Motor and sensory signals in vM1→vS1 axons. **a**, Image of the field of view and ROIs (red) corresponding to all visible axonal varicosities. Yellow numbers indicate ROI1 and ROI2. **b**, Activity ($\Delta F/F_0$) of two individual axons (ROI1 and ROI2) across three trials of one behavioural session, aligned with different behavioural features (whisker position, whisking set-point, whisking amplitude, $\Delta\kappa$, licking (ticks)). Protraction forces correspond to negative $\Delta\kappa$. Coloured bars indicate the sampling period (blue, go trials; red, no-go trials). **c**, Activity across an entire session for two axons ($\Delta F/F_0$ ROI1; $\Delta F/F_0$

ROI2). Each row corresponds to a trial. Dashed white lines indicate the sampling period when the pole is within reach. Trial type is indicated on the far left (hues of blue correspond to different object locations in go trials, as indicated in Fig. 1b). Only correct trials are shown. Right, behavioural variables (see Methods). $\Delta\kappa$ due to protractions; retractions are not shown. Whisking amplitude is the amplitude of the band-pass filtered (6–30 Hz) whisker angle. Lick rate is the instantaneous rate of tongue protractions. **d**, Activity (left) and behavioural variables (right) averaged across trial types. Shading represents s.e.m.

to quantify how well specific behavioural features could be decoded from axonal activity.

We first decoded the measured behavioural features ($\Delta\kappa$; whisking and licking; Methods) on the basis of activity in all axons in a field of view (Fig. 3a). The algorithms used the activity of populations of axons to fit individual behavioural features, taking into account variability within and across trials. The explained variance (R_i^2 , for the i th behavioural feature) was used to measure the quality of decoding (Supplementary Fig. 4).

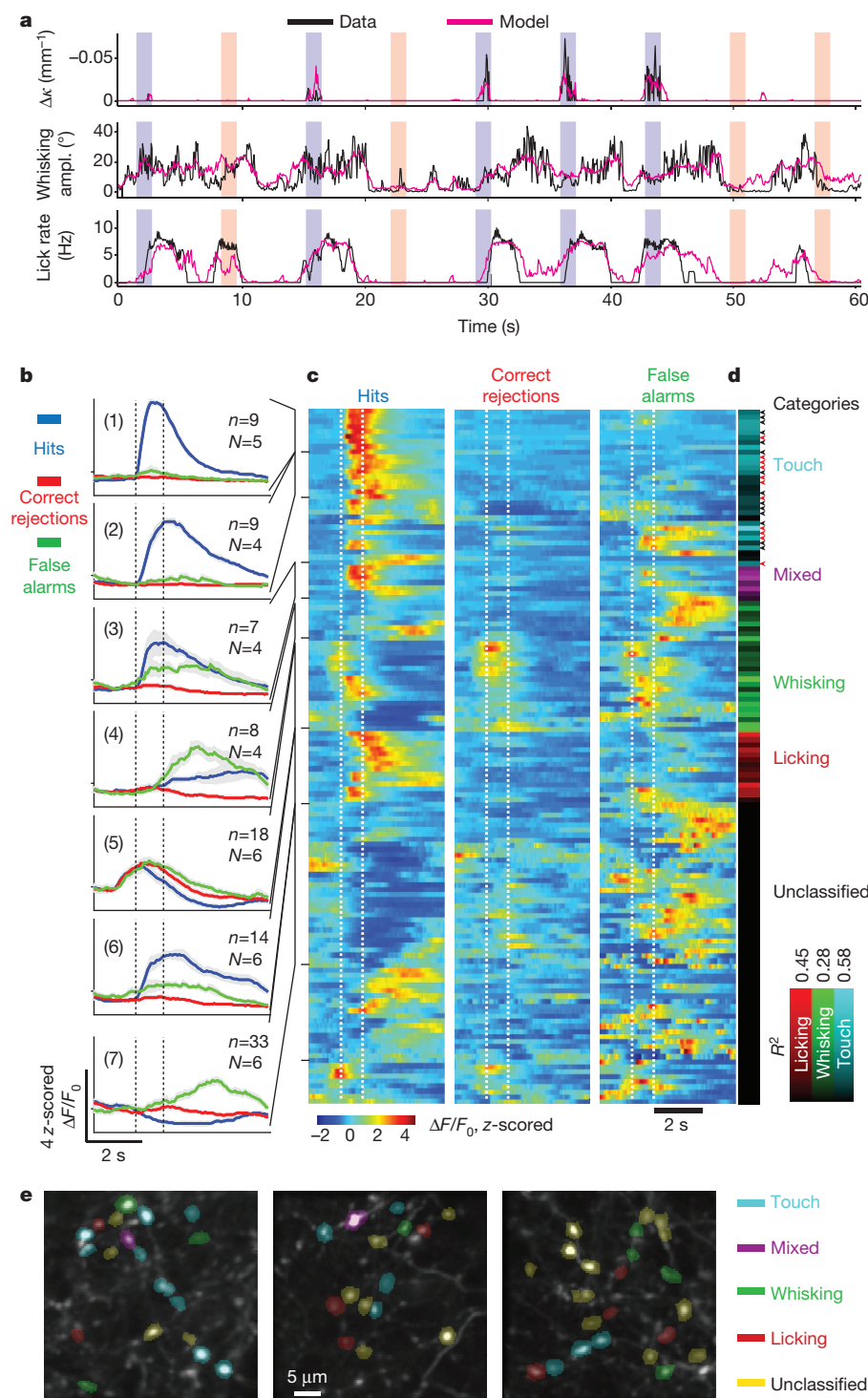
Population activity decoded the recorded behavioural features, including $\Delta\kappa$, whisking and lick rate (Fig. 3a). Whisking was decomposed into set-point (<1.6 Hz) and amplitude (6–30 Hz; Methods)^{6,12} (Fig. 2b), but these parameters were highly correlated (Supplementary

Fig. 5a) and decoding was similar for these parameters. All imaged fields of view decoded all behavioural features (R^2 larger than for trial-shuffled data; $P < 0.001$, bootstrap test for all sessions and features (Supplementary Fig. 4a)). This is despite the fact that each experiment sampled only approximately $1 \times 10^{-5}\%$ of the synapses in vS1 (assuming: thickness of L1 in vS1, 0.09 mm (ref. 22); vS1 area, 2 mm²; synaptic density, 1×10^9 per mm (refs 3 and 23)). We conclude that vM1→vS1 axons relay signals related to touch, whisking, licking and other task-related variables to vS1 in a distributed and highly redundant manner.

We used similar methods to classify individual axons on the basis of their activity patterns¹² (Methods; Fig. 3b–e). One half of the active axons (78 out of 138 active axons; 56%) decoded one or more of the

Figure 3 | Decoding behavioural variables on the basis of axonal activity. a, Decoding on the basis of axon populations. Time series of behavioural features down-sampled to 16 Hz (black) and the Random Forests model prediction (pink) based on the activity of all axons in one field of view. Vertical bars indicate the sampling periods and the trial type. Top, $\Delta\kappa$ induced by touch during protractions (whisker C2), a measure of contact force. Centre, whisking amplitude. Bottom, lick rate. b–e, Decoding behavioural features on the basis of the activity of individual axons. Axons were classified based on correlations with behavioural features (138 axons, 6 animals, 17 sessions).

b, Peristimulus time histograms averaged across axons with shared activity patterns and trial types, in standardized units (z-score). Hits (blue) are correct go trials, correct rejections (red) are correct no-go trials, and false alarms (green) are incorrect no-go trials. Incorrect go trials were rare and are not shown. Correct trials and error trials were used separately in classification. n , number of axons. 1, touch axons; 2, touch axons with persistent activity; 3, mixed axons; 4, whisking axons, late activity; 5, whisking axons, early activity; 6, licking axons; 7, axons that are selectively active during false alarms; these were unclassified using the behavioural features used. Grey shading represents s.e.m. c, Rows correspond to axons. Task-aligned activity was averaged over trials of each type. Rows were ordered on the basis of clustering by activity pattern (Methods). d, Correlation between the Random Forests model and different behavioural features (R_i^2) after classification (Methods). The hue indicates the strongest correlation with one of the features (cyan, touch; magenta, mixed; green, whisking; red, licking). Black and red arrowheads indicate axons showing object-location-dependent activity. Red arrowheads indicate significantly persistent axons (Fig. 4). e, Axons decoding different behavioural features were spatially intermingled (three sessions from three different animals). Left panel, same axons as in Fig. 2a; right panel, same axons as in Fig. 1f.



measured behavioural features (mean R^2 for best feature, 0.22). Whisking-related axons (26 out of 138 active axons) were active during whisking for all trial types, independent of touch (Fig. 2; ROI1). Whisking-related fluorescence signals lagged whisking onset (mean \pm s.d., 259 ± 101 ms) (Supplementary Fig. 5b–d). However, because of the slow dynamics of calcium-dependent fluorescence we cannot exclude the possibility that the underlying spikes might lead whisking. Coding of whisking amplitude and whisking set-point is consistent with neurophysiological^{6,24} and imaging experiments in vM1 (ref. 12). Touch-related axons (31 out of 138 active axons) were activated mainly in trials with touch, but not by whisking alone (Fig. 4b; Supplementary Fig. 4c). Consistently, trimming whiskers, which

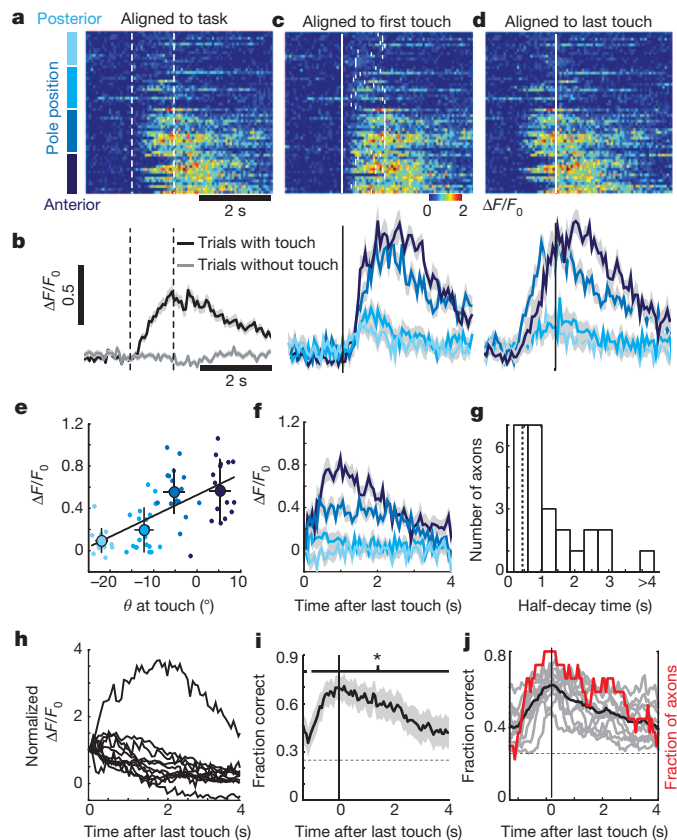


Figure 4 | Persistent object-location-dependent activity. **a**, Activity of an example axon for different object locations (hues of blue on the left; compare with Fig. 1b; only correct go trials are shown). Activity was aligned to the start of the trial (vertical dashed lines indicate presence of the pole). **b**, Activity averaged across trials with touch (black) and without touch (grey) aligned to start of trial. **c**, Same data as in **a**, aligned to the first touch (vertical line). Ticks, last touch in each trial. Bottom, activity averaged across object locations, aligned to first touch (hits). Grey shading indicates s.e.m. **d**, Same data as in **a**, aligned to the last touch. Bottom, activity averaged across object locations, aligned to last touch. **e**, Activity as a function of θ at touch. Same axon as in **a** is shown. For averaging, points were grouped by object location. Black line denotes linear regression (R^2 , 0.41; $P = 2.5 \times 10^{-9}$). Error bars denote s.d. **f**, Activity averaged across object locations, aligned to the last touch and deconvolved to correct for the dynamics of calcium and GCaMP3 fluorescence (same axon as in **a**) (see Supplementary Fig. 8). **g**, Decay time of fluorescence after last touch for all axons showing object-location-dependent activity ($n = 25$). The dashed vertical line indicates $T_{1/2}$ for GCaMP3 fluorescence without persistent activity. **h**, Average fluorescence signals for axons showing persistent activity ($n = 11$). Activity was normalized to the value at the time of the last touch. **i**, The fraction of trials with correctly decoded object location as a function of time for the example axon shown in **a–d**. Dashed line indicates chance level (0.25, corresponding to four object locations). Horizontal line indicates time of significant decoding ($*P < 0.05$). Grey shading indicates s.e.m. **j**, Decoding of object location (same group as in **h**; black line, average; red line, fraction of axons decoding above chance level).

precluded touch during whisking, abolished activity of touch-related axons (Supplementary Fig. 6). Other axons decoded lick rate (classified as lick related)²⁵ or a mixture of features (classified as ‘mixed’ axons) (Fig. 3d).

Axons representing different behavioural features were spatially intermingled over length scales of micrometres (Figs 2 and 3e); nearby varicosities were equally likely to be part of any of the representations. Repeated imaging experiments showed that the representations of individual axons were largely stable (Supplementary Fig. 6). The L1 apical tuft dendrites of individual neurons in vS1 may have access to a rich amalgam of contextual information.

Most of the touch-related axons showed activity that varied with object location (Fig. 3d, arrow heads; Fig. 4a–c and Supplementary Fig. 7) (25 out of 138 active axons). These axons became active only after first touch (Fig. 4a–c) (latencies greater than 100 ms) and might thus be driven in part by touch-triggered excitation spreading from vS1 to vM1 (refs 10, 11 and 13). In many cases, activation of these axons correlated with forces on a specific whisker. However, this whisker was typically not the principal whisker of the imaged barrel column (Supplementary Fig. 7). Touch-related activity corresponding to specific whiskers is thus broadcast widely to the barrel cortex, consistent with the diffuse nature of the vM1→vS1 projection⁴. These data implicate the primary motor cortex in somatosensation²⁶. Furthermore, signals coding for object location are fed back from the motor cortex to the somatosensory cortex.

Fluorescence changes coding for object location often outlasted the presence of the pole, sometimes for several seconds (Fig. 4d–h and Supplementary Fig. 8). This is much longer than expected for the decay-time constant of GCaMP3 fluorescence after trains of action potentials in axons (half-decay time, $T_{1/2}$, 0.45 s; Supplementary Fig. 2) (Fig. 4g, h). After taking calcium dynamics and GCaMP3 fluorescence into account, significant graded, persistent activity was seen in 11 out of 25 object-location-sensitive axons (two-way mixed analysis of variance (ANOVA), $P < 0.05$), lasting from 0.5 s to more than 4 s (Fig. 3d, red arrowheads; Fig. 4g, h). Motor behaviours did not explain object-location-dependent persistent activity; licking and whisking were not different across trials with different object locations (two-way mixed ANOVA, 9 out of 11 persistent object location axons, $P > 0.05$; Supplementary Fig. 8f, g). Object location could be decoded from the activity of persistent axons for up to 4 s after the last contact (Fig. 4i, j). This parametric persistent activity²⁶ represents a form of short-term memory of objects detected by specific whiskers at particular locations.

We used calcium imaging with genetically encoded indicators to track activity in specific projections (Fig. 1). Small patches of L1 in vS1 receive diverse signals from vM1, including activity coding for aspects of whisking, touch and licking (Figs 2 and 3). Most of the signals coded by L2/3 neurons in vM1 (ref. 12) were also represented in vM1 axons in L1 of the vS1. Touch-related and persistent activity was overrepresented in the vM1→vS1 axons, but this difference could be due to variations in the behavioural task or the sampled neuronal population.

Pyramidal neurons receive this rich top-down information in their tuft branches in L1, whereas bottom-up sensory input impinges mainly on the proximal basal dendrites⁵. L1 input increases neuronal gain and can promote bursting with coincident input in the proximal basal dendrites²⁷. Pyramidal neurons therefore compare contextual input and ongoing sensory input.

Activity related to whisker position was a prominent component of the vM1→vS1 signals (Figs 2 and 3). Pyramidal neurons in vS1 might combine this efference copy-like positional information and touch input to compute object location in vS1 (refs 6 and 15). Whisking signals impinging on L1 dendrites might also selectively amplify activity related to touch during periods of active exploration.

Although mice responded within approximately 100 ms after touch in the object-localization task, a subpopulation of vM1→vS1 axons showed persistent object-location signals that outlasted the decision of

the animals by seconds (Fig. 4). These memory traces, which might not be relevant to solve this particular task, represent information about touches of specific whiskers in the recent past. Haptic object recognition is constructed from localization of several related object features, based on multiple touches over time^{28,29} with multiple whiskers³⁰. The vM1→vS1 projection links the past state of specific whiskers (input in L1) with present sensory input from other whiskers, ascending into vS1 through the sensory thalamus. The vS1↔vM1 loop thus has the capacity to integrate haptic information across time and space, key components of object recognition and somatosensory navigation. Pyramidal cells in vS1 receive L1 inputs related to past touch, whereas their proximal basal dendrites receive input signalling present touch. These neurons could selectively burst after specific sequences of touches with different whiskers, within time windows defined by the mnemonic activity in L1. In this scheme burst firing of populations of vS1 neurons codes for complex object shape (Supplementary Fig. 9).

METHODS SUMMARY

Virus expressing GCaMP3 (rAAV-synapsin (SYN)-GCaMP3, serotype 2/1, University of Pennsylvania Gene Therapy Program Vector Core) was injected (20 nl, 10 nl min⁻¹, 350 µm deep) into the left vM1 of adult (>P60) C57BL/6 male mice and a circular craniotomy was made over the ipsilateral vS1. Animals were trained in a whisker-dependent object-localization task. vM1 axons were imaged in the vS1 14–23 days after virus injection. Images of the whiskers were acquired using a high-speed camera at 500 frames per second. The whisker position (azimuthal angle at the whisker base, θ) and whisker shape were tracked using automated whisker tracking^{8,14}. Licking was detected using an optical lickport.

Full Methods and any associated references are available in the online version of the paper.

Received 16 September 2011; accepted 12 June 2012.

Published online 26 August 2012.

- Felleman, D. J. & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
- Caulier, L. J. & Connors, B. W. Synaptic physiology of horizontal afferents to layer I in slices of rat S1 neocortex. *J. Neurosci.* **14**, 751–762 (1994).
- Caulier, L. Layer I of primary sensory neocortex: where top-down converges upon bottom-up. *Behav. Brain Res.* **71**, 163–170 (1995).
- Veinante, P. & Deschenes, M. Single-cell study of motor cortex projections to the barrel field in rats. *J. Comp. Neurol.* **464**, 98–103 (2003).
- Petreanu, L., Mao, T., Sternson, S. M. & Svoboda, K. The subcellular organization of neocortical excitatory connections. *Nature* **457**, 1142–1145 (2009).
- Hill, D. N., Curtis, J. C., Moore, J. D. & Kleinfeld, D. Primary motor cortex reports efferent control of vibrissa motion on multiple timescales. *Neuron* **72**, 344–356 (2011).
- Knutsen, P. M., Pietr, M. & Ahissar, E. Haptic object localization in the vibrissal system: behavior and performance. *J. Neurosci.* **26**, 8451–8464 (2006).
- O'Connor, D. H. *et al.* Vibrissa-based object localization in head-fixed mice. *J. Neurosci.* **30**, 1947–1967 (2010).
- O'Connor, D. H., Peron, S. P., Huber, D. & Svoboda, K. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron* **67**, 1048–1061 (2010).
- Ferezou, I. *et al.* Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. *Neuron* **56**, 907–923 (2007).
- Kleinfeld, D., Sachdev, R. N., Merchant, L. M., Jarvis, M. R. & Ebner, F. F. Adaptive filtering of vibrissa input in motor cortex of rat. *Neuron* **34**, 1021–1034 (2002).
- Huber, D. *et al.* Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484**, 473–478 (2012).
- Mao, T. *et al.* Long-range neuronal circuits underlying the interaction between sensory and motor cortex. *Neuron* **72**, 111–123 (2011).
- Clack, N. G. *et al.* Automated tracking of whiskers in videos of head fixed rodents. *PLoS Comput. Biol.* **8**, e1002591 (2012).
- Knutsen, P. M. & Ahissar, E. Orthogonal coding of object location. *Trends Neurosci.* **32**, 101–109 (2009).
- Birdwell, J. A. *et al.* Biomechanical models for radial distance determination by the rat vibrissal system. *J. Neurophysiol.* **98**, 2439–2455 (2007).
- Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
- Trachtenberg, J. T. *et al.* Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature* **420**, 788–794 (2002).
- Masino, S. A., Kwon, M. C., Dory, Y. & Frostig, R. D. Characterization of functional organization within rat barrel cortex using intrinsic signal optical imaging through a thinned skull. *Proc. Natl Acad. Sci. USA* **90**, 9998–10002 (1993).
- De Paola, V. *et al.* Cell type-specific structural plasticity of axonal branches and boutons in the adult neocortex. *Neuron* **49**, 861–875 (2006).
- Cox, C. L., Denk, W., Tank, D. W. & Svoboda, K. Action potentials reliably invade axonal arbors of rat neocortical neurons. *Proc. Natl Acad. Sci. USA* **97**, 9724–9728 (2000).
- Hooks, B. M. *et al.* Laminar analysis of excitatory local circuits in vibrissal motor and sensory cortical areas. *PLoS Biol.* **9**, e1000572 (2011).
- De Felipe, J., Marco, P., Fairen, A. & Jones, E. G. Inhibitory synaptogenesis in mouse somatosensory cortex. *Cereb. Cortex* **7**, 619–634 (1997).
- Carvell, G. E., Miller, S. A. & Simons, D. J. The relationship of vibrissal motor cortex unit activity to whisking in the awake rat. *Somatosens. Mot. Res.* **13**, 115–127 (1996).
- Komiyama, T. *et al.* Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature* **464**, 1182–1186 (2010).
- Hernández, A. *et al.* Decoding a perceptual decision process across cortex. *Neuron* **66**, 300–314 (2010).
- Larkum, M. E., Senn, W. & Luscher, H. R. Top-down dendritic input increases the gain of layer 5 pyramidal neurons. *Cereb. Cortex* **14**, 1059–1070 (2004).
- Anjum, F., Turmi, H., Mulder, P. G., van der Burg, J. & Brecht, M. Tactile guidance of prey capture in Etruscan shrews. *Proc. Natl Acad. Sci. USA* **103**, 16544–16549 (2006).
- Davidson, P. W. Haptic judgments of curvature by blind and sighted humans. *J. Exp. Psychol.* **93**, 43–55 (1972).
- Krupa, D. J., Matell, M. S., Brisben, A. J., Oliveira, L. M. & Nicolelis, M. A. Behavioral properties of the trigeminal somatosensory system in rats performing whisker-dependent tactile discriminations. *J. Neurosci.* **21**, 5752–5763 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Hooks, N. Li, Z. Guo, J. Magee and J. Dudman for comments on the manuscript, N. Clack, V. Iyer and J. Vogelstein for help with software and D. Flickinger for help with microscope design.

Author Contributions L.P. and K.S. conceived the study. L.P. performed the experiments. L.P., D.A.G. and K.S. analysed the data. D.A.G. and D.H.O. contributed software. D.H. and D.H.O. helped with behavioural and imaging experiments. N.-I.X. performed key pilot studies. L.T. and L.L. provided reagents. L.P., D.A.G. and K.S. wrote the paper with comments from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.S. (svobodak@janelia.hmi.org).

METHODS

Virus injection and chronic-window preparation. All procedures were approved by the Janelia Farm Research Campus Institutional Animal Care and Use Committee. Surgeries were conducted on adult ($>P60$) male C57BL/6 mice under isoflurane anaesthesia (1.5–2%). Other drugs reduced potential inflammation (Ketofen, 5 mg kg⁻¹, subcutaneously) and provided local (Marcaine, 0.5%, injected under the scalp) and general analgesia (buprenorphine, 0.1 mg kg⁻¹, intraperitoneally). To label motor cortex axons a small craniotomy was performed over the left vM1 (coordinates relative to bregma, anterior, 1.3 mm; lateral, 0.7 mm). Virus expressing GCaMP3 (rAAV-synapsin(SYN)-GCaMP3, serotype 2/1, University of Pennsylvania Gene Therapy Program Vector Core) was injected (20 nl, 10 nl min⁻¹, 350 μ m deep) with a custom-made volumetric injection system (based on a Narishige MO-10 manipulator)⁵. Glass pipettes (Drummond) were pulled and bevelled to a sharp tip (30 μ m outer diameter). Pipettes were back-filled with mineral oil and front-loaded with viral suspension immediately before injection. Compared to synthetic calcium indicators, expression of protein sensors ensures labelling of long-range axons and also allows imaging in trained mice across multiple behavioural sessions, separated by days to weeks. The virus infected neurons in layers 2–5, including most vS1-projecting neurons¹³ (Supplementary Fig. 2a). A custom-machined titanium frame was cemented to the skull with dental acrylic (Lang Dental).

For imaging, a circular craniotomy was made over the left barrel cortex (vS1; diameter, 1.5 mm; centre relative to bregma, lateral, 3.5 mm; anterior, -1.4 mm; left hemisphere, Fig. 1d). An imaging window was constructed from two layers of standard microscope coverglass (Fisher; thickness, 170–210 μ m), joined with an ultraviolet curable optical glue (NOR-61, Norland). A larger piece was attached to the bone and a smaller insert fit snugly into the craniotomy. The bone surrounding the craniotomy was thinned to allow for a flush fit between the insert and underlying dura. The window was cemented in place using dental acrylic (Lang Dental). After 3 days of recovery the C-row whiskers were mapped using intrinsic signal imaging⁹ (Fig. 1d). Mice were anaesthetized with isoflurane (1%) after injection of chlorprothixene (1 mg kg⁻¹)³¹. Images were acquired through a cranial window through a Leica MZ12.5 microscope under 630 nm illumination. Images of the vasculature over the same field of view were taken under 530 nm light-emitting diode illumination. Individual whiskers were moved using a piezoelectric bimorph (0.75 mm, 3 mm from the base; 10 Hz; 4 s, repeated every 20 s; total time, 10–30 min). Acquisition was performed with Ephus (<http://www.ephus.org>)³². Water restriction was started 1 day later.

Behaviour. The behavioural task was modified from previous studies^{8,12} (Fig. 1a, b and Supplementary Fig. 1). Behavioural training began after the mice had restricted access to water for at least 7 days (1 ml per day). In the initial session mice first learned to lick for water rewards (~40 rewards). Animals were then trained to lick only in trials in which a metal pole was within easy reach of the whiskers on the right side of their face. The pole was mounted on a pneumatic linear slider (Festo; SLS-10-30-P-A Mini slide; P/N 170496) and was moved rapidly (~0.25 s) in and out of reach of the whiskers under computer control. The pole was within reach of the whiskers for ~1 s (sampling period). Tongue movements were tracked using an optical lickport.

Because the pole was placed in one of five randomly selected locations, it was impossible for mice to position their whiskers to solve the task with passive whisker stimulation. However, consistent with previous studies⁸, mice tended to focus their whisking on the rewarded target locations. Passive stimulation by the moving pole thus occurred in some trials (23% of trials; 6 sessions, 5 animals). In general, mice whisked to contact the pole, but whisking varied across trials (Fig. 2b, c and Supplementary Fig. 1). Whisking strategies also varied across mice. After removal of the pole, during an answer period lasting ~1.5 s, licking was rewarded with a drop of water in go trials and punished with an extra inter-trial interval of 4 s in no-go trials. Licking was ignored in other behavioural epochs. Multiple pole positions in go trials were introduced at this early training stage. The range of go positions was initially relatively small (4.5 mm) and was gradually increased with training. The pole positions were (relative to the C2 follicle at rest) (Supplementary Fig. 1b): lateral, 6.7–9.1 mm. Go positions, centre of anterior–posterior range, 0.35–3.32 mm; anterior–posterior range, 5.6–7.0 mm; no-go positions, anterior–posterior location, 8.9–13.6 mm; lateral, 6.7–9.1 mm. Although all pole locations were within reach of the long posterior whiskers (C1, C2), mice did not always touch the most anterior (no-go) location⁸ (Supplementary Fig. 1e).

Once the animals were trained ($d' > 1$), all whiskers apart from C row were trimmed on the right side of the face. Subsequent behavioural sessions were performed while imaging under the two-photon microscope.

Imaging. Imaging was performed 14–23 days after virus injection with a custom-made microscope (design available at <http://research.janelia.org/Svoboda/>). GCaMP3 was excited using a Ti:sapphire laser (Mai Tai, Spectra Physics) tuned to $\lambda = 925$ nm. We used GaAsP photomultiplier tubes (10770PB-40, Hamamatsu)

and a $\times 16$ (0.8 NA) microscope objective (Nikon). The field of view was $40 \times 40 \mu\text{m}$ (128×256 pixels) (Fig. 1f), imaged at 16 Hz. The microscope was controlled with ScanImage³³ (<http://www.scanimage.org>). The average power for imaging was <50 mW, measured at the entrance pupil of the objective. For each mouse the optical axis was adjusted to be perpendicular to the imaging window. Imaging was stopped during inter-trial intervals (3 s; 7 s after false-alarm trials). Bleaching of GCaMP3 was negligible. Slow drifts of the field of view were corrected manually approximately every 50 trials using a reference image. In some cases we imaged the same axons over several days (Supplementary Fig. 6). Vascular landmarks and visual comparison with reference images from previous days were used to identify the same axons.

Calcium imaging has some drawbacks. Presently used genetically encoded indicators are not sufficiently sensitive to detect single action potentials, and as a consequence axons with low firing rates were probably missed^{19,17} (Supplementary Fig. 2). This implies that the representations in vM1→vS1 axons are probably even richer than suggested by our imaging data. In addition, the slow dynamics (hundreds of milliseconds) of the calcium indicator limits the temporal resolution of the neurophysiological measurements. Advances in the development of fluorescent sensors of neuronal function will probably ameliorate these problems.

Image analysis. Frames were registered using rigid translation based on cross-correlation³⁴. First, we selected a trial with little movement and the frames within the trial were registered and averaged. All of the frames from a session were registered to this averaged image. To extract fluorescence signals, ROIs were drawn over fluorescent varicosities identified by using the mean, maximum intensity and standard deviation values of all trials. The pixels in each ROI were averaged to estimate fluorescence corresponding to a single varicosity. The baseline fluorescence of the ROI, F_0 , was estimated as the 30th percentile of the fluorescence using a 32 s sliding window and used to calculate $\Delta F/F_0 = ((F - F_0)/F_0)$. To produce an event vector from the $\Delta F/F_0$ trace, and thereby minimize the temporal distortions caused by GCaMP3 dynamics¹⁷, we used a non-negative deconvolution method³⁵. These event vectors were used to train the decoder and to measure average activity when selecting for active axons. Active axons were defined as having an average event rate higher than 0.007 events per frame and a peak $\Delta F/F_0$ in the peristimulus time histograms higher than 30% for any trial type (that is, hit, correct rejection or false alarm). This assures that the calcium signals are not due to motion artefacts (less than 1% of the GCaMP3 signals are expected to be polluted by movement) (Supplementary Fig. 3).

Whisker tracking. Whiskers were illuminated using a high-power light-emitting diode (940 nm, Roithner) and condenser optics (Thorlabs). Images were acquired through a telecentric lens ($\times 0.36$, Edmund Optics) by a high-speed CMOS camera (EoSense CL, Mikrottron) running at 500 frames per second (640×352 pixels; 42 pixels per mm). Image acquisition was controlled by Streampix 3 (Norpix, Canada). The whisker position (θ) and whisker shape were tracked using automated whisker tracking^{8,14} (<https://openwiki.janelia.org/wiki/display/MyersLab/Whisker+Tracking>). Whiskers are cantilevered beams, with one end embedded into the follicle in the whisker pad. Measurement of whisker shape and the mechanical properties of the whisker can be used to estimate the forces acting on the follicle¹⁶. The amplitudes of the forces in the follicle are proportional to the curvature change of the whisker, $\Delta\kappa$. We used $\Delta\kappa$ at a particular location along the whisker (2–3 mm) as a substitute for the mechanical forces acting on the whiskers^{14,16}. Curvature was measured from a parametric curve comprising second-order polynomials fitted to the whisker backbone. Periods of contact between the whisker and object (touch) were detected on the basis of nearest distance between whisker and object.

Behavioural features. Licking was detected using a lickometer⁸. Lick rate (Hz) was the inverse of the inter-lick interval. Motor cortex neurons primarily code for slow variables related to whisking^{6,24}. We decomposed whisking (that is, θ at base) into whisking set-point and whisking amplitude. Whisking set-point was defined as the 0.6 s moving average of θ . Whisker amplitude was defined as the Hilbert transform³⁶ of the absolute value of the band-pass filtered (6–30 Hz) θ . Whisking set-point and whisker amplitude were highly correlated in our behavioural task (Supplementary Fig. 5a) and only one of these features (whisking amplitude) is shown in most figures (Figs 2 and 3). To characterize the quality of the decomposition we reconstructed whisker position from set-point, whisker amplitude and whisking phase. The reconstruction error (root mean squared error) across all animals was $4.5 \pm 0.18^\circ$ (mean \pm s.e.m.; range 3.5 – 6.1°), corresponding to 30% of the variance of whisker position. The analysis of curvature was restricted to periods of touch. We further derived features related to protraction touch (negative $\Delta\kappa$) and retraction touch (positive $\Delta\kappa$) and separately absolute values.

All behavioural features were down-sampled to the image-acquisition rate (16 Hz). Mean and maximum values were calculated for each feature in a 64-ms window centred on the middle of the new sampling point. Mice whisk at frequencies above 8 Hz (the Nyquist frequency)³⁷ and this information is lost by

down-sampling. Across animals, $48 \pm 1\%$ of the energy of the whisking trajectory was contained at frequencies below 8 Hz. The higher frequency signals (that is, phase) are not thought to be coded by most of the motor cortex neurons^{6,24}. Also, our imaging methods would be too slow to capture modulation with whisking phase. **Correlation analysis to determine if multiple varicosities are part of the same axon.** vM1 axons form elaborate arborizations in L1 of the vS1, and multiple varicosities in one field of view were often part of the same axon. To characterize the diversity of signals impinging on vS1, we wanted to report activity in distinct axons, as opposed to distinct varicosities. We used correlation-based methods to distinguish varicosities that were part of the same axon and those that were part of different axons (Fig. 1h–j).

We first selected varicosities that, on the basis of structural images, were part of the same axon (14 sessions). We computed their correlation coefficient over an entire session, and compared them to correlation coefficients computed over all pairs of varicosities (dominated by pairs from different axons). As expected^{21,38}, pairs of varicosities from the same axon showed markedly higher correlation coefficients compared with random pairs. The correlation coefficient increases with increasing event rates. For varicosities from the same axon we plotted the correlation coefficients as a function of event rates and fitted an exponential curve to the data (Fig. 1j). The 95% confidence value of the fit was used as a criterion to assign pairs of varicosities to the same axon, even in the cases where the axonal morphology could not be discerned.

To build clusters of correlated ROIs we selected all the pairs that were considered to be from the same axon on the basis of correlations (Fig. 1h–j). We seeded a cluster with one randomly selected pair. The next randomly selected pair could share one of the ROIs with the existing cluster, in which case it joined the cluster; otherwise it seeded a second cluster, and so on. We iterated this procedure until all pairs were assigned. Each cluster was represented in the data set by a 'representative' ROI, defined by the largest mean $\Delta F/F_0$.

Characterizing movement using mice with GFP-expressing axons. Axons and their varicosities are tiny structures (one micrometer diameter, or less²⁰). Movement of the animal could displace axons in and out of the objective focal plane and thus produce motion artefacts in the fluorescence signal. To quantify these errors we performed control experiments in three mice with vM1→vS1 axons labelled with green fluorescent protein (GFP). All experimental conditions were identical, except that rAAV-SYN-GFP (serotype 2/1) was injected into vM1 instead of the GCaMP3-expressing virus. These mice were trained in the object-localization behaviour and analysed identically to mice expressing GCaMP3 (Supplementary Fig. 3). Although varicosities in GFP animals showed movement-induced changes in fluorescence, mainly during the licking period, these changes rarely exceeded 25%.

Characterization of axonal varicosities in brain slices. Male C57BL/6 mice (P15) were injected in the left vM1 with 20 nl AAV-SYN-GCaMP3, serotype 2/1 as described⁵. Two weeks later vS1 slices were prepared. Mice were anaesthetized with an intraperitoneal injection of a ketamine/xylazine mixture (0.13 mg ketamine and 0.01 mg xylazine per gram of body weight) and perfused through the heart with ice-cold artificial cerebrospinal fluid (ACSF; ~5 ml) containing 127 mM NaCl, 25 mM NaHCO₃, 25 mM D-glucose, 2.5 mM KCl, 1 mM MgCl₂, 2 mM CaCl₂ and 1.25 mM NaH₂PO₄, aerated with 95% O₂ and 5% CO₂. The brain was removed and placed into ice-cold cutting solution containing 110 mM choline chloride, 25 mM NaHCO₃, 25 mM D-glucose, 11.6 mM sodium ascorbate, 7 mM MgCl₂, 3.1 mM sodium pyruvate, 2.5 mM KCl, 1.25 mM NaH₂PO₄ and 0.5 mM CaCl₂. Coronal slices (300 µm thick) of the right barrel cortex were cut with a vibrating slicer (Microm, Walldorf) and incubated in oxygenated ACSF for 45 min at 37 °C and then at room temperature (approximately 23 °C). For recording, slices were placed in a chamber with circulating ACSF at 34 °C under a custom-built two-photon microscope. A 40 × 40 µm area of L1 in the barrel cortex was imaged at 16 Hz using a Ti:sapphire laser tuned to $\lambda = 925$ nm. L1 axons were stimulated with an extracellular electrode at 83 Hz. Fluorescence traces were extracted as described for the *in vivo* experiments.

Decoding behavioural variables. The relationship between fluorescence signal x_i of the i th axon and the j th behavioural variable y_j can be characterized as an encoding description $P(x_i|y_j)$ or a decoding description $P(y_j|x_i)$. The encoding description specifies how much of the activity of the axon can be accounted for by the behavioural variables. The decoding description specifies how well a behavioural variable can be derived from the activity of populations of axons (Fig. 3a and Supplementary Fig. 4) or single axons (Figs 3b–e and 4i, j). Here, we focused on the decoding description.

We used machine learning algorithms to decode the behavioural variables based on activity. For single axon decoding (Fig. 3b–e) each axon was used to predict behavioural variables, such as forces on each whisker, whisking amplitude and lick rate. The input to the algorithm was the event rate (that is, deconvoluted $\Delta F/F_0$) of a given axon as well as time-shifted versions. To predict sensory input we used present

and future activity. For motor variables we used both past and future values, as neural activity could reflect motor commands, corollary discharges or re-afferent input.

The goal of the decoder algorithm was to find a mapping $\hat{y}_j(t_k) = f[x_i(t_{k-l}), \dots, x_i(t_k), \dots, x_i(t_{k+p})]$ that best approximates $y_j(t_k)$ for all t_k . t_k is discretized time (in units of 1/16 s, corresponding to the imaging rate); l and p represent the maximum negative and positive shifts of the activity, respectively.

We concatenated trials to generate a vector \bar{t} of time-binned data. For sensory variables we used $l = 5$ and $p = 0$ and for sensory-motor variables $l = 5$ and $p = 5$ (corresponding to time-shifts up to 0.32 s). The dimensionality of the input variables is $l + p + 1$.

For decoding neural populations (Fig. 3a and Supplementary Fig. 4a) we used for each session all axons showing at least one event and created an input vector of size $N_{\text{axons}} \times (l + p + 1)$. Each behavioural session, corresponding to one imaged field of view, was treated separately.

The algorithm was trained on a subset of trials (the training set, 80%) and evaluated on a separate set of test trials (20%). We repeated this procedure five times to obtain a prediction for all trials³⁹.

The accuracy of the decoding algorithm was evaluated using the Pearson correlation coefficient (ρ) between the model estimate and the data. The explained variance is $R^2 = \rho^2$ (range, 0–1). R^2 was calculated separately for each trial type (that is, hit, correct rejection, miss and false alarm). Treating trial types separately was critical to disambiguate the relationship between different behavioural variables and activity of single axons. For instance, we observed large-amplitude whisking during licking, which complicates the classification of the response type of the axon. However, during correct-rejection trials, licking was absent and whisking present, simplifying the classification of axon types. Similarly, in trained animals, touch and licking occurred with short latencies in hit trials. By contrast, in false-alarm trials touch was typically absent or diminished in magnitude.

Decoding was with Random Forests^{39,40}, a multivariate nonparametric machine learning algorithm based on bootstrap aggregation (that is, bagging) of regression trees. We used the TreeBagger class implemented in Matlab. TreeBagger requires only few parameters: the number of trees ($N_{\text{trees}} = 32$), the minimum leaf size (minleaf = 10), the number of features chosen randomly at each split ($N_{\text{split}} = N_{\text{features}}/3$; the typical value used by default⁴⁰). These parameters were chosen as a trade-off between decoder accuracy and computation time. We did not observe much improvement in decoding accuracy for $N_{\text{trees}} > 32$ and minleaf < 10 (data not shown).

Classification of response types. We measured the R^2 between each measured behavioural variable (that is, whisking amplitude, lick rate, whisking set-point, etc) and each axon's decoder prediction for all the trials and for each trial type (Fig. 3b–e). In addition, we used ANOVA to determine whether calcium responses were different for different pole locations (Fig. 4). We grouped the behavioural variables in larger categories such as whisking (that is, including whisking amplitude and whisker set-point), lick rate and touch (that is, touch per whisker, rate of change of forces, absolute magnitude, etc). We considered the best R^2 set for each of the three behavioural categories. Alternatively, all axons were classified manually on the basis of trial-to-trial calcium transients and behavioural prediction for each trial type. For most axons classification was unambiguous based on the decoder R^2 values. The remaining axons were more accurately classified based on a rarer trial type. For example, touch-related axons were much more active in trials with touches compared to trials without touches (Supplementary Fig. 4c). Lick-related axons, but not touch-related axons, were active during licking in false-alarm trials. Three of the authors independently arrived at consistent classifications.

Clustering of response types. We used clustering algorithms to display the trial-averaged responses of all active unique axons (Fig. 3b). First, we divided the axons into five categories: touch, whisking, licking, mixed and unclassified. Within each category we computed a similarity index between pairs of axons based on the trial-averaged responses in hit, correct-rejection and false-alarm trials. For each axon we concatenated their trial-averaged response for these three trial types. The similarity index was the pairwise correlation between the concatenated trial-average responses between every axonal pair. We computed a weighted correlation to take into account that the number of false-alarm trials is typically lower than either hit or correct-rejection trials.

Given the two vectors \bar{y} and \bar{x} and the weight vector \bar{w} (by the relative number of trials in each trial type):

$$\text{Weighted mean: } \mu(\bar{x}; \bar{w}) = \frac{\sum_i w_i x_i}{\sum_i w_i}$$

$$\text{Weighted covariance: } \text{cov}(\bar{x}, \bar{y}; \bar{w}) = \frac{\sum_i w_i (x_i - \mu(\bar{x}; \bar{w}))(y_i - \mu(\bar{y}; \bar{w}))}{\sum_i w_i}$$

$$\text{Weighted correlation: } \text{corr}(\bar{x}, \bar{y}; \bar{w}) = \frac{\text{cov}(\bar{x}, \bar{y}; \bar{w})}{\sqrt{\text{cov}(\bar{x}, \bar{x}; \bar{w}) \text{cov}(\bar{y}, \bar{y}; \bar{w})}}$$

Axons were reordered using hierarchical clustering for each category separately (using linkage and dendrogram from Matlab with unweighted average distances between clusters).

Analysis of sustained activity. For all cells showing object-location-dependent activity we tested for parametric persistent activity. Fluorescence traces were deconvolved with an exponential decay characterized by the GCaMP3 decay-time constant ($T_{1/2} = 0.45$ s) (compare Fig. 4d, f and Supplementary Fig. 8c, d). The mean value of the deconvolved traces at different times after contact (0.2–1.1, 1.1–2.0, 2.0–2.9 and 2.9–3.8 s after the last contact) was tested for object-location-dependent activity (two-way mixed-effect ANOVA). To determine whether object-location-dependent motor behaviours could explain object-location-dependent activity, similar analyses were performed for lick rate and whisking parameters (Supplementary Fig. 8f, g).

We applied linear and nonlinear decoders (Random Forests, linear and quadratic discriminant analysis; naive Bayes was also tested and gave identical results to linear discriminant analysis) to determine the time course of the object-location memory traces and compared their performance (Supplementary Fig. 8h). Each decoder was trained to predict the pole position as a function of time on the basis of the average activity (that is, events) in the previous 0.5 s for each axon. We performed 200 repetitions of this procedure to compute the average performance of each decoder and their confidence intervals. We ensured that there was an equal number of trials per each of the four possible pole positions (that is, chance level corresponded to 0.25). As before, we used 80% of the trials to train the decoders and tested the performance in the remaining 20% trials (repeating this five times to obtain a prediction for the full set of trials). The linear and quadratic discriminant decoders assume that the

input is a multivariate Gaussian and find the optimal linear or quadratic curve that best separates the classes⁴¹.

31. Kaneko, M., Hanover, J. L., England, P. M. & Stryker, M. P. TrkB kinase is required for recovery, but not loss, of cortical responses following monocular deprivation. *Nature Neurosci.* **11**, 497–504 (2008).
32. Suter, B. A. *et al.* Ephus: multipurpose data acquisition software for neuroscience experiments. *Front. Neural Circuits* **4**, 100 (2010).
33. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser-scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
34. Guizar-Sicairos, M., Thurman, S. T. & Fienup, J. R. Efficient subpixel image registration algorithms. *Opt. Lett.* **33**, 156–158 (2008).
35. Vogelstein, J. T. *et al.* Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **104**, 3691–3704 (2010).
36. Hill, D. N., Bermejo, R., Zeigler, H. P. & Kleinfeld, D. Biomechanics of the vibrissa motor plant in rat: rhythmic whisking consists of triphasic neuromuscular activity. *J. Neurosci.* **28**, 3438–3455 (2008).
37. Voigts, J., Sakmann, B. & Celikel, T. Unsupervised whisker tracking in unrestrained behaving animals. *J. Neurophysiol.* **100**, 504–515 (2008).
38. Koester, H. J. & Sakmann, B. Calcium dynamics associated with action potentials in single nerve terminals of pyramidal cells in layer 2/3 of the young rat neocortex. *J. Physiol.* **529**, 625–646 (2000).
39. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning* 2nd edn (Springer, 2009).
40. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
41. Duda, R. O., Hart, P. E. & Stork, D. G. *Pattern Classification* 2nd edn (Wiley, 2001).

CORRIGENDUM

doi:10.1038/nature11372

Corrigendum: Recurrent network activity drives striatal synaptogenesis

Yevgenia Kozorovitskiy, Arpiar Saunders, Caroline A. Johnson, Bradford B. Lowell & Bernardo L. Sabatini

Nature **485**, 646–650 (2012); doi:10.1038/nature11052

On page 3 of the PDF and print version of this Letter, column two, in the text describing the mEPSC frequency and dendritic spine density values shown in Fig. 3b and 3c, the words ‘saline’ and ‘cno’ were erroneously exchanged. The correct text should read: (D1-Cre, direct pathway MSNs: frequency, saline 0.8 ± 0.1 Hz, $n = 18$; cno 0.36 ± 0.07 Hz, $n = 25$; spine density, saline 0.96 ± 0.04 spines per μm , $n = 5$; cno 0.49 ± 0.05 spines per μm , $n = 5$; D2-Cre, indirect pathway MSNs: frequency, saline 0.04 ± 0.01 Hz, $n = 7$; cno, 0.27 ± 0.09 Hz, $n = 8$; spine density, saline 0.54 ± 0.05 spines per μm , $n = 5$; cno 0.98 ± 0.06 spines per μm , $n = 5$). This has been corrected online in the PDF and HTML of the original paper.

CORRIGENDUM

doi:10.1038/nature11373

Corrigendum: Biodiversity loss and its impact on humanity

Bradley J. Cardinale, J. Emmett Duffy, Andrew Gonzalez,
David U. Hooper, Charles Perrings, Patrick Venail,
Anita Narwani, Georgina M. Mace, David Tilman,
David A. Wardle, Ann P. Kinzig, Gretchen C. Daily,
Michel Loreau, James B. Grace, Anne Larigauderie,
Diane S. Srivastava & Shahid Naeem

Nature **486**, 59–67 (2012); doi:10.1038/nature11148

In Table 1 and Supplementary Table 2 of this Review, under the ‘Category of service’ called ‘Regulating’, the first two ‘Measures of service provision’ related to ‘Biocontrol’ should read ‘Abundance of herbivorous pests’ instead of ‘Control of herbivorous pests’. With this word change, a downward arrow for either the predicted or actual diversity–service relationship would indicate that the abundance of herbivorous pests declines (and biocontrol increases) with increasing plant diversity. This does not alter any of our conclusions, because all diversity–service relationships were correctly described in the text of the manuscript itself. These errors have been corrected online in the HTML and PDF versions of the original Review, and in the original Supplementary Information.

Observations of increased tropical rainfall preceded by air passage over forests

D. V. Spracklen¹, S. R. Arnold¹ & C. M. Taylor²

Vegetation affects precipitation patterns by mediating moisture, energy and trace-gas fluxes between the surface and atmosphere¹. When forests are replaced by pasture or crops, evapotranspiration of moisture from soil and vegetation is often diminished, leading to reduced atmospheric humidity and potentially suppressing precipitation^{2,3}. Climate models predict that large-scale tropical deforestation causes reduced regional precipitation^{4–10}, although the magnitude of the effect is model^{9,11} and resolution⁸ dependent. In contrast, observational studies have linked deforestation to increased precipitation locally^{12–14} but have been unable to explore the impact of large-scale deforestation. Here we use satellite remote-sensing data of tropical precipitation and vegetation, combined with simulated atmospheric transport patterns, to assess the pan-tropical effect of forests on tropical rainfall. We find that for more than 60 per cent of the tropical land surface (latitudes 30 degrees south to 30 degrees north), air that has passed over extensive vegetation in the preceding few days produces at least twice as much rain as air that has passed over little vegetation. We demonstrate that this empirical correlation is consistent with evapotranspiration maintaining atmospheric moisture in air that

passes over extensive vegetation. We combine these empirical relationships with current trends of Amazonian deforestation to estimate reductions of 12 and 21 per cent in wet-season and dry-season precipitation respectively across the Amazon basin by 2050, due to less-efficient moisture recycling. Our observation-based results complement similar estimates from climate models^{4–10}, in which the physical mechanisms and feedbacks at work could be explored in more detail.

To explore the links between vegetation and rainfall, we analysed combined satellite data on precipitation from the Tropical Rainfall Measuring Mission (TRMM) and other satellites¹⁵ (TRMM3B42) and data on leaf area index (LAI) from the Moderate Resolution Imaging Spectroradiometer¹⁶ (MODIS). In these data positive spatial correlations exist between annual mean precipitation and annual mean LAI (Fig. 1a, d; Pearson correlation coefficient, $r = 0.81$), highlighting the role of precipitation in controlling large-scale vegetation patterns. In this study, our aim was to investigate a causal effect of vegetation on tropical (30° S to 30° N) rainfall in subsequent days on a regional scale (over distances of hundreds to thousands of kilometres). To do this, we calculated the origin and atmospheric transport of air

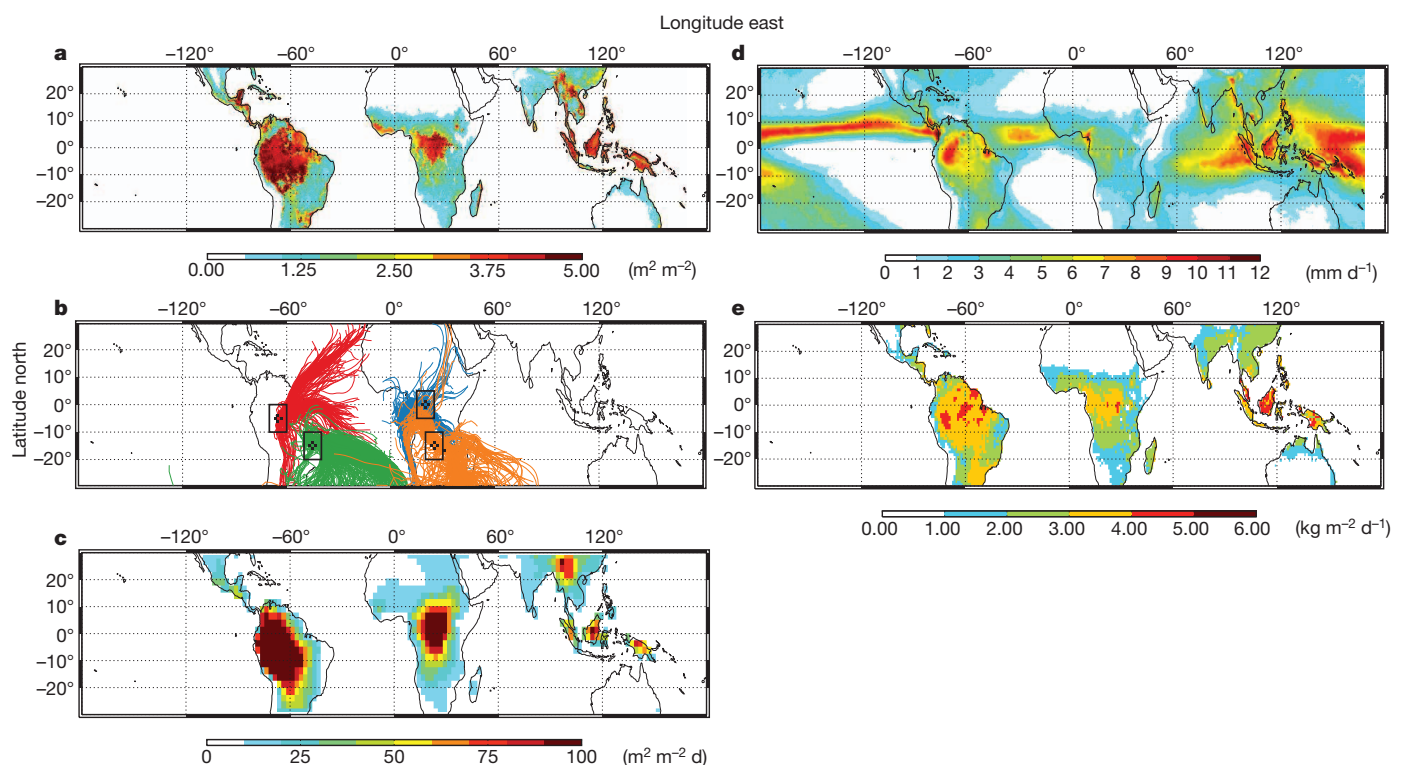


Figure 1 | Annual (2001–2007) mean vegetation, precipitation and evaporation. **a**, Leaf area index (LAI) from MODIS. **b**, Example 10-d back-trajectories arriving daily during 2001. Boxes illustrate the four domains analysed in detail in this study. **c**, Mean cumulative exposure of back-trajectories to LAI over the preceding 10 d. **d**, Precipitation reported by TRMM and other satellites (TRMM3B42). **e**, Evapotranspiration computed as the mean of the four GLDAS models.

¹School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK. ²Centre for Ecology and Hydrology, Wallingford, Oxford OX10 8BB, UK.

masses to determine the prior exposure of air to vegetation. This allowed us to evaluate whether air that had passed over more vegetation produced more rainfall.

To calculate air-mass histories, we used a Lagrangian atmospheric transport model. Such back-trajectory methods have been used previously to identify the transport of atmospheric moisture to continental regions^{17,18}. We calculated atmospheric back-trajectories arriving daily at the centre of all continental $1^\circ \times 1^\circ$ grid squares over the tropical domain for 2001–2007. The trajectories were calculated using operational analysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF), and are hence consistent with the large-scale atmospheric flow from the assimilated observations. We tested the sensitivity of our analysis to the length, arrival height and arrival time of the back-trajectories and found consistent results across a broad range of choices (Supplementary Figs 1, 2 and 3). Here we present results for ten-day back-trajectories arriving at the surface at 12:00 UT. Figure 1b shows example back-trajectories arriving at four tropical locations.

For each individual trajectory, we calculated the cumulative LAI (Σ LAI) encountered by the air mass during the last ten days of atmospheric transport. The resulting climatology of Σ LAI (Fig. 1c) is very similar to the *in situ* LAI, although important differences are apparent. For example, there are regions surrounding the Amazon and Congo basins where *in situ* LAI is relatively low but Σ LAI is high owing to exposure of air to large amounts of vegetation as it travels across forested regions upwind.

We analysed relationships between the daily variability in Σ LAI and the daily variability in precipitation. Figure 2a shows these relationships for air masses arriving in a $\sim 1,000 \text{ km} \times 1,000 \text{ km}$ region of Minas Gerais, Brazil, near the Amazon basin. To reduce the influence of the initial state of the air mass on our analysis, we stratified the data into that from the wet season and that from the dry season and then further according to the initial specific humidity of the trajectory, on the basis of the ECMWF analyses. We found a strong positive and significant (Student's *t*-test, $P < 0.01$) relationship between the exposure of air masses to vegetation and the precipitation those air masses produce (Fig. 2b). For air masses with low to medium exposure to precedent vegetation (Σ LAI $< 10 \text{ m}^2 \text{ m}^{-2} \text{ d}$), mean dry- and wet-season rainfall increases by 0.25 mm d^{-1} and, respectively, 0.4 mm d^{-1} for every additional unit of exposure to LAI that is encountered in the preceding 10 d. The impact of vegetation exposure tends to saturate, especially during the wet season, with less sensitivity of rainfall where Σ LAI $> 10 \text{ m}^2 \text{ m}^{-2} \text{ d}$. Although we find that the initial specific humidity of the air mass does affect precipitation, with moister air masses typically producing more precipitation, the positive trend between precipitation and exposure to LAI is similar for all subsets of the data. Qualitatively similar results are found in a $1,000 \text{ km} \times 1,000 \text{ km}$ region south of the tropical forests of the Congo basin (Fig. 2b). Over the moist tropical forests themselves, differences in rainfall between air masses with low and high exposures to LAI are notably smaller (Fig. 2b), although for the Congo basin the differences are significant in the dry season ($P < 0.01$). We note that the positive relationships found between precipitation and Σ LAI are not due to the length of time the trajectories have spent over land, with distance travelled and Σ LAI often poorly correlated (Supplementary Table 1), or to the average topographic height the trajectory has crossed. When we repeated our analysis using the distance travelled by the trajectory over land or the average elevation of the topography crossed by the trajectory (in place of Σ LAI), the relationships were substantially weaker and in some cases negative (Supplementary Figs 4 and 5).

We extended this regional analysis to explore relationships between precipitation and vegetation across the tropics (Fig. 2c). For more than 60% of the tropical land surface, precipitation is a factor of at least two greater in air masses that have been exposed to extensive vegetation in the preceding days (defined as the top decile of Σ LAI), relative to air masses that have been exposed to little (defined as the bottom decile of

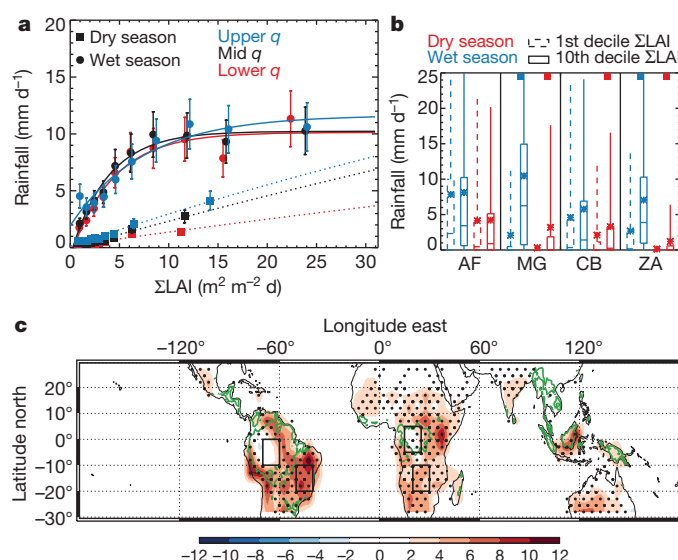


Figure 2 | Relationships between daily precipitation and cumulative exposure of 10-d back-trajectories to vegetation LAI (Σ LAI) for 2001–2007. **a**, Plot for air masses arriving in Minas Gerais, Brazil ($10\text{--}20^\circ \text{ S}$, $40\text{--}50^\circ \text{ W}$). Data binned into deciles of Σ LAI and stratified by initial specific humidity (q). Lines show fit to data (solid, wet season; dotted, dry season) and error bars indicate estimation of error in precipitation (Methods Summary). **b**, Comparison of daily precipitation for air masses that have been exposed to small and large amounts of vegetation (significant ($P < 0.01$) differences indicated by squares at top of panel) during atmospheric transport to the Amazon basin (AB; $10\text{--}0^\circ \text{ S}$, $60\text{--}70^\circ \text{ W}$), Minas Gerais (MG), the Congo basin (CB; $5^\circ \text{ N--}5^\circ \text{ S}$, $15\text{--}25^\circ \text{ E}$) and south of Congo (ZA; $10\text{--}20^\circ \text{ S}$, $20\text{--}30^\circ \text{ E}$) (mean, star; median, line; 25th and 75th percentiles, box; 5th and 95th percentiles, whiskers). **c**, Number of calendar months with significant ($P < 0.01$; red, positive; blue, negative) relationships between precipitation and Σ LAI. Stippling denotes regions where precipitation is a factor of at least two greater in air with large exposure to vegetation than in air with small exposure. Green contour delimits areas with $> 3 \text{ m}^2 \text{ m}^{-2}$ annual mean LAI. Black boxes mark the four regions in **b**.

Σ LAI). Significant ($P < 0.01$) positive correlations between precipitation and Σ LAI are common features for much of the year in areas surrounding the Amazon (southern Brazil and Paraguay) and Congo (southern and eastern Africa) forests, matching where previous studies have found large continental precipitation recycling ratios¹⁸. We find few tropical regions with significant negative correlations, although the relationships between vegetation and precipitation are typically weaker in moist tropical forests. The weaker signal at the centre of extensive forests is probably due to the lack of variability in air-mass exposure to vegetation (Supplementary Fig. 6); however, saturation of the MODIS LAI retrievals for dense tropical forest canopies¹⁹ may have a role.

This analysis demonstrates that there are strong positive relationships between the cumulative exposure of air to vegetation and the amount of precipitation that air will produce, suggestive of a water-cycle feedback. To explore potential mechanisms underlying these relationships, we evaluated the change in atmospheric moisture that occurs along our back-trajectories. To do this we calculated the net change in specific humidity (Δq) that occurred during continental transport in the ECMWF humidity analyses along the 10-d back-trajectories (Fig. 3a). In general, air becomes drier during atmospheric transport over land, owing to lower continental evaporation rates as compared with the oceans. Figure 3a demonstrates that air masses that have been exposed to more vegetation remain significantly moister ($P < 0.01$), and in some cases air can actually moisten when crossing densely vegetated regions. Analysis of the latter cases indicates that typically 70–90% of increases in q occur during the hours of daylight,

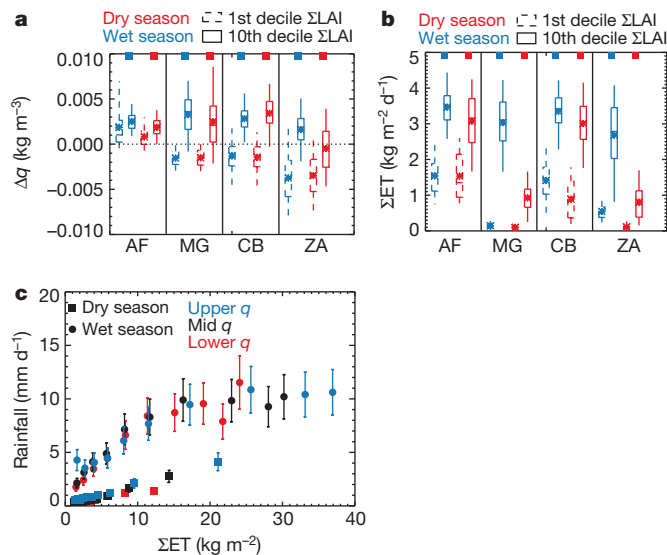


Figure 3 | Atmospheric water-budget components along back-trajectories. **a**, Same as Fig. 2b, but for net change in atmospheric specific humidity (Δq) as a function of ΣLAI . **b**, Same as Fig. 2b, but for cumulative surface evaporation (ΣET) as a function of ΣLAI . **c**, Same as Fig. 2a, but for precipitation as a function of ΣET .

consistent with a dominant forcing from evapotranspiration (Supplementary Fig. 7).

To quantify the contribution of evapotranspiration to the air-mass water budget, we used output from four global land surface models that use the best available estimates of meteorological forcing (those from the Global Land Data Assimilation System²⁰ (GLDAS)) (Fig. 1e). Figure 3b shows significant ($P < 0.01$) positive relationships between the multi-model mean cumulative surface evaporation (ΣET) and ΣLAI in all four regions examined. Figure 3c shows rainfall as a function of ΣET for the Minas Gerais region, demonstrating that additional moisture from evapotranspiration emitted into air masses with large exposure to vegetation is substantially greater than the additional precipitation observed in these air masses. Indeed, for all four regions the extra ΣET emitted into air masses with large vegetation exposure exceeds the observed additional precipitation by a factor of at least four (Supplementary Table 2).

Our analysis explores the role of regional-scale vegetation patterns on precipitation. Through evapotranspiration, forests maintain atmospheric moisture that can return to land as rainfall downwind. These processes operate on timescales of days over distances of 100–1,000 km (ref. 18) such that large-scale land-use change may alter precipitation hundreds to thousands of kilometres from the region of vegetation change. Land-use patterns and small-scale deforestation may also alter precipitation locally, through changes in the thermodynamic profile and the development of surface-induced mesoscale circulations^{21,22}. Natural and pyrogenic emissions from vegetation can also have a role in rainfall initiation over tropical forest regions²³. The impact of cloud microphysical processes on precipitation is highly uncertain²⁴, and biogenic emissions could contribute to our observed relationship between rainfall and exposed vegetation. However, our water-balance calculations imply that cumulative increases in evapotranspiration over upstream forested regions more than account for the increase in downstream rainfall.

Rapid land-use change is occurring across large regions of the tropics: 40% of the Amazon is predicted to be deforested by 2050 under a business-as-usual scenario²⁵. We used this scenario to explore the potential sensitivity of rainfall to changes in moisture recycling as a result of deforestation. We combined the deforestation scenario with present-day LAI to produce a new spatial distribution of LAI

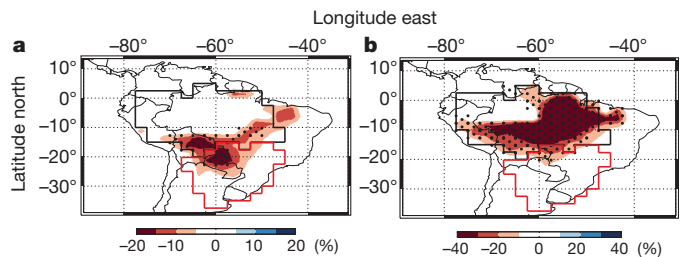


Figure 4 | Simulated percentage change in precipitation due to 2000–2050 business-as-usual deforestation of the Amazon basin. **a**, Wet season; **b**, dry season. Stippling denotes regions where the simulated precipitation anomaly differs from the present-day (1998–2010) rainfall by more than 1 s.d. The Amazon (black) and Rio de la Plata (red) basins are marked.

(Supplementary Fig. 8) and then used our trajectories to calculate ΣLAI under the deforested scenario. We then applied our empirical relationships between ΣLAI and rainfall (Fig. 2b) to estimate the change in rainfall that might occur as a result of this extensive deforestation. In this calculation, we implicitly assumed there to be no change in the large-scale circulation and that the local effects of deforestation on rainfall were negligible downstream. Despite these assumptions, our estimates are broadly consistent with estimates of basin-wide deforestation from climate models^{4–10}. Using this method, we estimated a 12% reduction in wet-season precipitation and a 21% reduction in dry-season precipitation across the Amazon basin (Fig. 4). This sensitivity is not restricted to the region of deforestation, and we estimate a 4% decrease in annual total precipitation for the Rio de la Plata basin. Through comparison with TRMM3B42 data, we calculate that the estimated reduction in precipitation is equivalent to the basin-wide drought experienced across the Amazon in 2010. Such a reduction in precipitation may have consequences for the future of remaining Amazonian forests^{26,27} and for rainfall-reliant industries both within and outside the Amazon basin, including agriculture and hydroelectric power generation, which contribute substantially to South American economies. The successful efforts to curb Amazon deforestation that have been applied in recent years²⁸ must be maintained if large-scale clearance of the Amazon and the resulting impacts on regional rainfall are to be avoided.

METHODS SUMMARY

Remote-sensed data. We used precipitation retrievals from the 3B42 3-h $0.25^\circ \times 0.25^\circ$ product of TRMM and other satellites¹⁵ to calculate daily accumulated (24-h) rainfall. We used monthly mean LAI from MODIS¹⁶ using the MOD15_BU_V5 product available at $0.25^\circ \times 0.25^\circ$ resolution. We spatially averaged both products to $1^\circ \times 1^\circ$ resolution. We apply this temporal and spatial averaging to the precipitation data to reduce random error²⁹. We estimate the total error to be the random error, estimated according to ref. 28, plus a systematic error estimated as 0.2 times the absolute precipitation³⁰, combined in quadrature.

Land surface models. We used $1^\circ \times 1^\circ$ monthly mean evaporation from four global land surface models archived on GLDAS²⁰. The models are forced by a combination of meteorological data sets including atmospheric analysis, and precipitation from merged gauge–satellite products.

Atmospheric transport. We calculated kinematic atmospheric back-trajectories arriving four times daily (00:00, 06:00, 12:00 and 18:00 UT) on a $1^\circ \times 1^\circ$ grid for the period 2001–2007 using the OFFLINE trajectory model. The position of each trajectory is calculated every 30 min and output every 6 h. We calculated 3-, 5- and 10-d trajectories arriving at the surface and three altitudes above the surface (corresponding to air pressures of 900, 800 and 700 hPa) that are likely to be within the deep tropical boundary layer.

Analysis. For each trajectory, we calculated the total distance travelled over land, the cumulative exposure to LAI (ΣLAI), the average elevation of the topography over which the trajectory travelled and the cumulative evapotranspiration (ΣET ; specified by GLDAS). We fitted rainfall data with a function of the form $f(x) = ae^{bx} + c$ (Supplementary Table 2). We calculated ΣLAI regardless of trajectory pressure. Restricting the calculation to when trajectory pressure is greater than 850 hPa gives similar results (Supplementary Fig. 9).

Full Methods and any associated references are available in the online version of the paper.

Received 16 April; accepted 29 June 2012.

Published online 5 September; corrected online 12 September 2012 (see full-text HTML version for details).

1. Bonan, G. B. Forests and climate change: forcings, feedbacks, and the climate benefit of forests. *Science* **320**, 1444–1449 (2008).
2. Shukla, J. & Mintz, Y. Influence of land-surface evapotranspiration on the Earth's climate. *Science* **215**, 1498–1501 (1982).
3. Eltahir, E. A. B. & Bras, R. L. Precipitation recycling in the Amazon basin. *Q. J. R. Meteorol. Soc.* **120**, 861–880 (1994).
4. Henderson-Sellers, A. & Gornitz, V. Possible climatic impacts of land cover transformations, with particular emphasis on tropical deforestation. *Clim. Change* **6**, 231–257 (1984).
5. Lean, J. & Warrilow, D. A. Simulation of the regional climatic impact of Amazon deforestation. *Nature* **342**, 411–413 (1989).
6. Shukla, J., Nobre, C. & Sellers, P. Amazon deforestation and climate change. *Science* **247**, 1322–1325 (1990).
7. Werth, D. & Avissar, R. The local and global effects of Amazon deforestation. *J. Geophys. Res.* **107**, 8087 (2002).
8. Ramos da Silva, R., Werth, D. & Avissar, R. Regional impacts of future land-cover changes on the Amazon basin wet-season climate. *J. Clim.* **21**, 1153–1170 (2008).
9. Hasler, N., Werth, D. & Avissar, R. Effects of tropical deforestation on global hydroclimate: a multimodel ensemble analysis. *J. Clim.* **22**, 1124–1141 (2009).
10. Nobre, P. *et al.* Amazon deforestation and climate change in a coupled model simulation. *J. Clim.* **22**, 5686–5697 (2009).
11. Pitman, A. J. *et al.* Uncertainties in climate responses to past land cover change: first results from the LUCID intercomparison study. *Geophys. Res. Lett.* **36**, L14814 (2009).
12. Butt, N., de Oliveira, P. A. & Costa, M. H. Evidence that deforestation affects the onset of the rainy season in Rondonia, Brazil. *J. Geophys. Res.* **116**, D11120 (2011).
13. Negri, A. J., Adler, R. F., Xu, L. & Surratt, J. The impact of Amazonian deforestation on dry season rainfall. *J. Clim.* **17**, 1306–1319 (2004).
14. Chagnon, F. J. F. & Bras, R. L. Contemporary climate change in the Amazon. *Geophys. Res. Lett.* **32**, L13703 (2005).
15. Huffman, G. *et al.* The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **8**, 38–55 (2007).
16. Myneni, R. B. *et al.* Global products of vegetation leaf area index and fraction absorbed PAR from year one of MODIS data. *Remote Sens. Environ.* **83**, 214–231 (2002).
17. Gimeno, L., Drumond, A., Nieto, R., Trigo, R. M. & Stohl, A. On the origin of continental precipitation. *Geophys. Res. Lett.* **37**, L13804 (2010).
18. van der Ent, R. J., Savenije, H. H. G., Schaefli, B. & Steele-Dunne, S. C. Origin and fate of atmospheric moisture over continents. *Wat. Resour. Res.* **46**, W09525 (2010).
19. Aragão, L. E. O. C. *et al.* Spatial validation of collection 4 MODIS LAI product in eastern Amazonia. *IEEE Trans. Geosci. Rem. Sens.* **43**, 2526–2534 (2005).
20. Rodell, M. *et al.* The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
21. Wang, J. F., Bras, R. L. & Eltahir, E. A. B. The impact of observed deforestation on the mesoscale distribution on the mesoscale distribution of rainfall and clouds in Amazonia. *J. Hydrometeorol.* **1**, 267–286 (2000).
22. Garcia-Carreras, L. & Parker, D. J. How does local tropical deforestation affect rainfall? *Geophys. Res. Lett.* **38**, L19802 (2011).
23. Andreae, M. O. *et al.* Smoking rain clouds over the Amazon. *Science* **303**, 1337–1342 (2004).
24. Koren, I. *et al.* Aerosol-induced intensification of rain from the tropics to mid-latitudes. *Nature Geosci.* **5**, 118–122 (2012).
25. Soares-Filho, B. S. *et al.* Modelling conservation in the Amazon basin. *Nature* **440**, 520–523 (2006).
26. Phillips, O. L. *et al.* Drought sensitivity of the Amazon rainforest. *Science* **323**, 1344–1347 (2009).
27. Aragão, L. E. O. C. *et al.* Spatial patterns and fire response of recent Amazonian droughts. *Geophys. Res. Lett.* **34**, L07701 (2007).
28. Davidson, E. A. *et al.* The Amazon basin in transition. *Nature* **481**, 321–328 (2012).
29. Huffman, G. J. Estimates of root-mean-square random error for finite samples of estimated precipitation. *J. Appl. Meteorol.* **36**, 1191–1201 (1997).
30. Smith, T. M., Arkin, P. A., Bates, J. J. & Huffman, G. J. Estimating bias of satellite precipitation estimates. *J. Hydrometeorol.* **7**, 841–856 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements D.V.S. acknowledges a Natural Environment Research Council grant (NE/G015015/1). The GLDAS data used in this study were acquired as part of the mission of NASA's Earth Science Division and were archived and distributed by the Goddard Earth Sciences Data and Information Services Center.

Author Contributions D.V.S. and S.R.A. initiated the project. All authors participated in discussions, conducted the analysis, assisted with data interpretation and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.V.S. (d.v.spracklen@leeds.ac.uk).

METHODS

Remote-sensed data. We used precipitation retrievals (from the Tropical Rainfall Measuring Mission (TRMM; <http://trmm.gsfc.nasa.gov/>) and other satellites) in the 3B42 3-h $0.25^\circ \times 0.25^\circ$ product¹⁵ to calculate daily accumulated (24 h) rainfall. This product reports precipitation from microwave sensors on board polar-orbiting satellites, combined with more frequent cloud-top temperature data, surface rain gauge analysis and the precipitation radar on TRMM. We used the monthly mean leaf area index (LAI) from the Moderate Resolution Imaging Spectroradiometer (MODIS; http://modis.gsfc.nasa.gov/data/dataproducts.php?MOD_NUMBER=15)¹⁶ using the MOD15_BU_V5 product available at $0.25^\circ \times 0.25^\circ$ resolution.

The precipitation product includes both systematic and random error^{29,30}. Systematic biases of $0.5\text{--}1\text{ mm d}^{-1}$, or 0.2 times the mean precipitation, have been estimated³⁰. To reduce the random error we spatially averaged the precipitation product to $1^\circ \times 1^\circ$ resolution and temporally averaged to give daily accumulated (24 h) precipitation. This temporal and spatial averaging applied to TRMM3B42 means that each data point in our analysis (which corresponds to a single trajectory) is the average of 128 TRMM3B42 data points. This substantially reduces the random error that is present in the TRMM3B42 product²⁹. We estimate the random error using the method of ref. 29 and assume a systematic error of 0.2 times the mean precipitation³⁰. We combine the random and systematic errors in quadrature to give an estimate of the total error, which in any case is substantially smaller than the large effect of vegetation on precipitation. We spatially averaged the LAI data to the same resolution as the precipitation product.

Land surface models. We used $1^\circ \times 1^\circ$ monthly mean evaporation from four global land surface models (Community Land Model (CLM), Variable Infiltration Capacity model (VIC), NOAH and MOSIAC) archived on the Global Land Data Assimilation System (GLDAS; <http://disc.sci.gsfc.nasa.gov/services/grads-gds/gldas>)²⁰. The models are forced by a combination of meteorological data sets including atmospheric analysis, and precipitation from merged gauge–satellite products.

Atmospheric transport. We calculated kinematic atmospheric back-trajectories arriving daily (00:00, 06:00, 12:00 and 18:00 UT) on a $1^\circ \times 1^\circ$ grid for the period 2001–07 using the OFFLINE trajectory model³¹. The position of each trajectory is calculated every 30 min, and output every 6 h. We calculated 3-, 5- and 10-day trajectories arriving at the surface and 3 levels above the surface (900, 800 and 700 hPa) that are likely to be within the deep tropical boundary layer. We demonstrate that our results are robust to the back-trajectory length (Supplementary Fig. 1), arrival pressure (Supplementary Fig. 2) and arrival time (Supplementary Fig. 3) of the trajectory. Throughout the Letter we report analysis of 10-day back-trajectories arriving at the surface at 12:00 UT.

Analysis. For each trajectory we calculated (1) total distance travelled over land (Σdist), (2) cumulative exposure to LAI (ΣLAI), (3) average elevation of the topography (specified by the Climate Research Unit CRU CL 2.0 database, $10'$ resolution) over which the air mass travels, and (4) cumulative evapotranspiration (specified by GLDAS; ΣET). We calculate ΣLAI regardless of trajectory pressure but show that restricting the calculation to when trajectory pressure is greater than 850 hPa gives similar results (Supplementary Fig. 9).

We analysed relationships between daily precipitation and the variables calculated from air mass history (Σdist , ΣLAI and ΣET). We stratified trajectories into dry season and wet season and according to the initial specific humidity (q) of the back-trajectory (taken from the ECMWF analyses). The timings of wet and dry season are location dependent. For the purpose of this analysis we defined the dry season at any location as calendar months with below annual average precipitation for that location and the wet season as calendar months with above average precipitation (as observed by TRMM3B42; Supplementary Table 2). We demonstrated that this stratification results in large variability in initial q of the back-trajectory (Supplementary Table 2). We conducted detailed analysis over 4 large ($10^\circ \times 10^\circ$, $\sim 1,000\text{ km} \times 1,000\text{ km}$) domains. Our pan-tropical analysis was conducted at a horizontal resolution of $2.5^\circ \times 2.5^\circ$. For our 7-year analysis each $10^\circ \times 10^\circ$ grid box represents 255,000 trajectories whereas each $2.5^\circ \times 2.5^\circ$ grid box represents 15,968 trajectories.

We binned trajectories into deciles of ΣLAI . We then compared back-trajectories with small exposure to vegetation (lowest decile of ΣLAI) to back-trajectories with large exposure to vegetation (largest decile of ΣLAI) and used the Student's t -test to determine the significance of any differences. Accounting for the stratification described above means that each data point (for example, data point in Fig. 2a, box-plot in Fig. 2b) represents 3,650 trajectories. We fitted relationships between precipitation and ΣLAI with functions of the form $y(x) = a\exp(bx) + c$ (see Fig. 2a). Supplementary Table 2 gives the fitted variables for the $10^\circ \times 10^\circ$ domains. Using a linear fit ($y(x) = ax + c$), which does not capture the nonlinear behaviour of the data well, altered our estimated impacts of deforestation on annual mean Amazon basin rainfall from -14% to -10% .

To estimate the impact of Amazonian deforestation on precipitation we combined the functions we fitted above with projected LAI distributions after deforestation. We created an LAI distribution for the year 2050 (see Supplementary Fig. 8) by combining a business-as-usual deforestation scenario²⁵ with the present-day LAI distribution from MODIS. We assumed that deforested areas are maintained as pasture with a LAI of $1\text{ m}^2\text{ m}^{-2}$ (ref. 32). We then ran our present-day trajectories over the projected LAI to calculate the ΣLAI that would occur in the deforestation scenario. We used our empirical relationships (calculated at a resolution of $2.5^\circ \times 2.5^\circ$) along with ΣLAI to estimate the rainfall that would occur after deforestation. We estimated both wet season and dry season rainfall. Our approach only estimates the change in rainfall due to changes in water recycling. It makes the implicit assumption that there is no change in the large-scale circulation, and that the local impacts of deforestation on rainfall are negligible downstream. Calculation of the full impacts of deforestation on rainfall would require a climate model.

We compared the estimated changes in rainfall due to deforestation with present-day (1998–2010) rainfall recorded by TRMM3B42. We make comparisons for both the wet season and the dry season.

31. Methven, J. *Offline Trajectories: Calculation and Accuracy* Technical Report 44, (UK Universities Global Atmospheric Modelling Programme, University of Reading, 1997).
32. Aragão, L. E. O. C., Shimabukuro, Y. E., Santo, F. D. B. E. & Williams, M. Landscape pattern and spatial variability of leaf area index in Eastern Amazonia. *For. Ecol. Mgmt.* **211**, 240–256 (2005).

Distinct contribution of stem and progenitor cells to epidermal maintenance

Guilhem Mascré¹, Sophie Dekoninck¹, Benjamin Drogat¹, Khalil Kass Youssef¹, Sylvain Brohée^{1,2}, Panagiota A. Sotiropoulou¹, Benjamin D. Simons^{3,4} & Cédric Blanpain^{1,5}

The skin interfollicular epidermis (IFE) is the first barrier against the external environment and its maintenance is critical for survival. Two seemingly opposite theories have been proposed to explain IFE homeostasis. One posits that IFE is maintained by long-lived slow-cycling stem cells that give rise to transit-amplifying cell progeny, whereas the other suggests that homeostasis is achieved by a single committed progenitor population that balances stochastic fate. Here we probe the cellular heterogeneity within the IFE using two different inducible Cre recombinase–oestrogen receptor constructs targeting IFE progenitors in mice. Quantitative analysis of clonal fate data and proliferation dynamics demonstrate the existence of two distinct proliferative cell compartments arranged in a hierarchy involving slow-cycling stem cells and committed progenitor cells. After wounding, only stem cells contribute substantially to the repair and long-term regeneration of the tissue, whereas committed progenitor cells make a limited contribution.

Skin epidermis is comprised of a basal layer of proliferative cells and several suprabasal layers of terminally differentiated cells that are progressively enucleated, forming squames that are shed from the skin surface¹. Depending on anatomical location and species, the minimum transit time from the basal layer to the cornified layer has been estimated at around 1–2 weeks². In homeostasis, cells lost during the course of turnover must be perfectly compensated by cells generated in the basal layer. Different theories have been proposed to explain how this balance is achieved¹. On the basis of morphological and proliferation studies, it has been proposed that the IFE is organized into discrete ‘epidermal proliferative units’ (EPU), comprised of slow-cycling stem cells (SCs) together with around 10 transit-amplifying cell progeny, which undergo terminal differentiation after a fixed number of cell divisions^{2–5}. Following the clonal marking of IFE cells by retroviruses^{6–9} or mutagens^{10,11}, long-lived columns of labelled IFE cells that span the epidermis from the basal layer to the top of the cornified layer appear, lending support to the concept of EPU. By contrast, the quantitative analysis of lineage tracing data in IFE using a ubiquitous promoter suggests that tissue is maintained by a single, equipotent, committed progenitor (CP) cell population in which the balance between proliferation and differentiation follows from seemingly random cell fate decisions^{12,13}. Although this model is attractive in its simplicity, it does not easily explain the ability of tissue to respond rapidly to increased cell demand, such as during wound healing, and cannot rule out the presence of a quiescent or slow-cycling IFE population¹⁴. More importantly, whereas the contribution of different populations of hair follicle SCs to wound healing is well-established^{15–18}, little is known about the contribution of putative IFE SCs or CP cells to the repair of skin epidermis.

From a quantitative analysis of lineage tracing data using two different Cre recombinase–oestrogen receptor (Cre-ER) transgenic mice, we demonstrate the existence of two distinct populations of epidermal progenitors that form a SC/CP cell hierarchy that differentially contribute to the homeostasis and repair of the epidermis.

Heterogeneity of epidermal progenitors

To probe the proliferative heterogeneity within the IFE, we made use of two Cre-ER transgenic mice that target IFE progenitors (Supplementary Fig. 1a, b). With a high dose of tamoxifen, Cre-ER under the control of the keratin 14 promoter (K14-Cre-ER) can target most cells in the basal layer of tail IFE¹⁹. A very low dose of tamoxifen (0.2 mg) to K14-Cre-ER/RosaYFP induced yellow fluorescent protein (YFP) expression at clonal density (Supplementary Fig. 1c). Although Cre-ER under the control of the involucrin promoter (Inv-Cre-ER) predominantly targets cells in the suprabasal layers, we have recently shown that some basal cells, characterized by K5 but not Inv protein expression, are also induced but at a lower frequency than K14-Cre-ER (Supplementary Fig. 1d–f)²⁰. The basal cells marked by these two different Cre-ER comprise IFE progenitors, as demonstrated by their ability to proliferate, and give rise to differentiated progeny, forming columns of labelled cells (Supplementary Fig. 1c, d).

To determine whether the two induction protocols mark the same homogenous pool of progenitors, we first quantified their survival at different time points after tamoxifen administration. Although many of the clones induced by the K14-Cre-ER survived up to 1 year post-induction, most of the Inv-Cre-ER targeted clones were progressively lost (Fig. 1a–c). Whereas the population of ‘surviving clones’, defined as clones that retain at least one basal layer cell, drops by around 30% for K14 mice from 4 weeks to 6 months after labelling, the clone density falls by a factor of 10 in the Inv mice over the same period, indicating that IFE is maintained by a heterogeneous pool of progenitors with different survival potential.

Inv-Cre-ER targets committed progenitors

To gain further insight into the behaviour of these apparently distinct pools of progenitors, we quantified the pattern of growth of individual clones targeted by the Inv-Cre-ER at eight time points ranging from 3.5 days to 48 weeks post-induction, a period spanning more than half of the average lifetime of the mouse. At 3.5 days, 40% of surviving clones had undergone at least one round of cell division (Fig. 2a, b and

¹Université Libre de Bruxelles, IRIBHM, Brussels B-1070, Belgium. ²Université Libre de Bruxelles, Machine Learning Group, Brussels B-1050, Belgium. ³Cavendish Laboratory, Department of Physics, J. J. Thomson Avenue, Cambridge CB3 0HE, UK. ⁴The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK. ⁵WELBIO, Université Libre de Bruxelles, Brussels B-1070, Belgium.

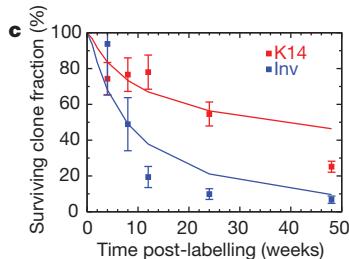


Figure 1 | K14-Cre-ER and Inv-Cre-ER target IFE progenitors with different survival characteristics. **a, b,** Immunostaining of K5 and YFP in K14-Cre-ER/RosaYFP (**a**) and Inv-Cre-ER/RosaYFP (**b**) tail epidermis at 1 week and 48 weeks post-induction showing the high survival rate of K14-Cre-ER clones. Dashed lines represent the basal lamina. Hoechst nuclear staining is represented in blue. **c,** Quantification of surviving clones in K14-Cre-ER/RosaYFP and Inv-Cre-ER/RosaYFP tail epidermis at different times following tamoxifen administration ($n = 327$ clones at least, for each time point of K14 and Inv). Symbols show experimental data and the lines represent the prediction of the mathematical modelling. The data has been normalized to the initial clone density as inferred by the fit to theory. Error bars show s.d. Scale bars, 200 μm .

Supplementary Fig. 2a), and the vast majority of these divisions resulted in one basal and one suprabasal cell (Fig. 2c, d). At 4 weeks post-induction, although almost all surviving clones had undergone at least one cell division, 40% of clones still contained only one basal cell (Fig. 2a, e), indicating that most of these divisions led to asymmetric fate outcome. However, the other 60% of surviving clones had also expanded basally (Fig. 2a, f), indicating that progenitors can also symmetrically self-renew.

This clonal heterogeneity increased progressively over the 48-week time course, indicating that progenitors follow divergent fates (Supplementary Fig. 2b, c). At the same time, the average size of surviving clones (as measured both by their total size and their ‘footprint’ on the basal layer) grew approximately linearly with time (Supplementary Fig. 2d), whereas the number of surviving clones progressively fell (Fig. 1c). Such behaviour mirrors that reported previously¹² in studies using a ubiquitous promoter, and indicates that Inv-Cre-ER targets the actively cycling CP cell population, in which stochastic cell loss through differentiation is perfectly compensated by duplication (Fig. 2g). This conclusion is reinforced by a quantitative analysis of the data, which confirms that the basal layer clone size distribution converges onto the hallmark scaling behaviour in which the chance of finding a clone larger than some multiple of the average

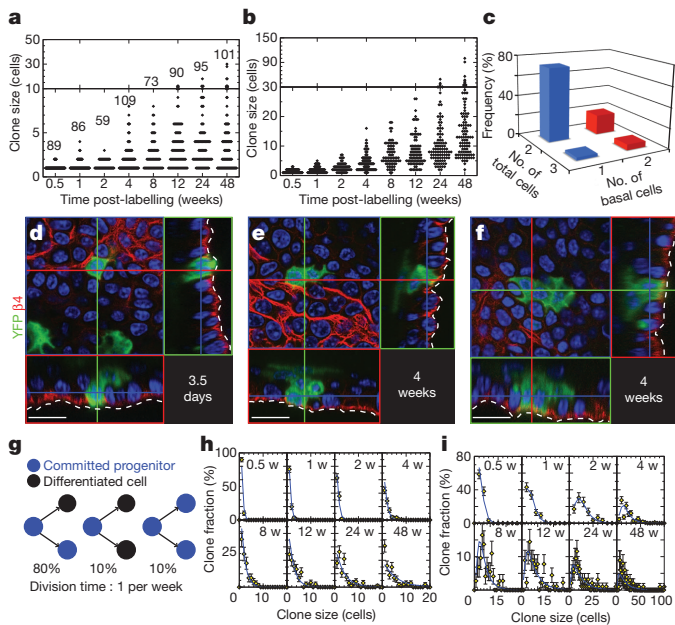


Figure 2 | Inv-Cre-ER targets IFE CP cells. **a, b**, Distribution of Inv-Cre-ER clone sizes as measured by basal (**a**) and total (**b**) cell content of surviving clones, imaged by confocal microscopy on whole-mount tail epidermis from 3.5 days to 48 weeks following tamoxifen administration. The number of analysed clones is indicated in **a** and this is identical in **b**, **h** and **i**. **c**, Size distribution of surviving clones with more than one cell at 3.5 days post-labelling disaggregated according to basal and total cell number ($n = 37$ clones). **d–f**, Confocal analysis of representative Inv-Cre-ER-targeted basal clones at 3.5 days (**d**) and 4 weeks (**e, f**) post tamoxifen. Dashed lines represent the basal lamina. Hoechst nuclear staining is represented in blue. **g**, A fit of the Inv-Cre-ER clonal fate data predicts that progenitors undergo population asymmetry through a combination of asymmetric cell division, symmetrical self-renewal and symmetrical differentiation with an average cell cycle time of approximately 1 week. **h, i**, Frequency distribution of basal (**h**) and total (**i**) clone size. Symbols represent experimental data and the lines correspond to the model prediction. Error bars show s.e.m. Scale bars, 10 μm .

stays constant over time (Supplementary Fig. 2e and Supplementary Methods, section ‘Theory’)²¹. From the detailed study of the basal and total clone size distribution over the 48 week time course (Fig. 2h, i, Supplementary Fig. 2f and Supplementary Methods, section ‘Theory’), we are able to conclude that, following CP cell division (at an average rate of 1.21 ± 0.04 per week), 80% result in asymmetric fate (leading to one dividing and one differentiated cell) with the remainder leading to symmetric duplication or differentiation with approximately equal probability (Fig. 2g), consistent with the results from ref. 12. Indeed, from the 48 week time point, we are able to resolve a small (less than 1 in 14) bias towards terminal division.

K14-Cre-ER targets long-lived SCs

To determine whether the dynamics of K14 clones are different from the Inv clones, we performed extensive clonal analysis using the K14-Cre-ER mice ranging from 3.5 days to 48 weeks post-induction (Fig. 3). At just 3.5 days, 60% of K14 clones had already undergone one round of cell division (Fig. 3a, b and Supplementary Fig. 3a), whereas the fraction of clones involving two or more basal cells was greatly increased compared with Inv (Fig. 3a-d). However, following an initial abrupt expansion, the growth of K14 clones rapidly decelerates over the first few weeks, with the size distribution from 8 weeks to 6 months post-induction showing only a modest expansion (Fig. 3a-f and Supplementary Fig. 3a-c). Alongside the propensity of clones to persist, these observations are consistent with the K14 promoter targeting both the CP cell population and a second, slow-cycling, SC population. In particular, the precipitous expansion of clones (Fig. 3a-c), which is far more rapid than that found in the

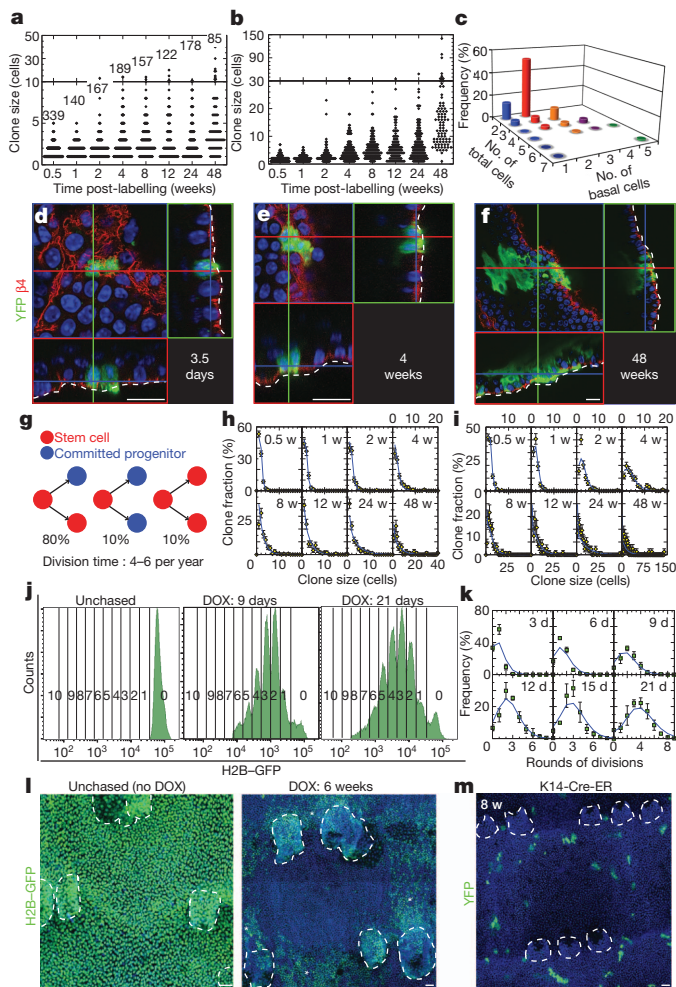


Figure 3 | K14-Cre-ER targets IFE SCs. **a, b**, Distribution of K14-Cre-ER clone sizes as measured by basal (**a**) and total (**b**) cell content of surviving clones, imaged by confocal microscopy on whole-mount tail epidermis from 3.5 days to 48 weeks following tamoxifen administration. The number of analysed clones is indicated in **a** and this is identical in **b**, **h** and **i**. **c**, Size distribution of surviving clones with more than one cell at 3.5 days post-labelling disaggregated according to basal and total cell number ($n = 197$ clones). **d–f**, Confocal analysis of representative K14-Cre-ER targeted basal clones at 3.5 days (**d**), 4 weeks (**e**) and 48 weeks (**f**) post tamoxifen. Dashed lines represent the basal lamina. Hoechst nuclear staining is represented in blue. Scale bars, 10 μm . **g**, Biophysical modelling of the clonal fate data predicts that K14-Cre-ER targets both CP cells and SCs with the latter undergoing population asymmetry through a combination of asymmetric cell division, symmetrical self-renewal and symmetric differentiation with an average cell cycle rate of between 4 and 6 per year. **h, i**, Frequency distribution of basal (**h**) and total (**i**) clone size. Symbols represent data and the lines correspond to the model prediction. **j**, FACS analysis of K5tTA/tet(O)-H2BGFP mice at different time point following DOX administration. **k**, Frequency of the number of cell divisions that surviving basal IFE cells undergo during the chase period as quantified by FACS analysis at different time points after DOX administration to adult K5tTA/tet(O)-H2BGFP mice. Symbols represent the average of the experimental data ($n = 3$ mice at each time point). Lines correspond to the predicted model dynamics (Supplementary Methods, section “Theory”). **l, m**, Confocal analysis of H2B-GFP immunostaining in unchased mice (no DOX) (left panel) and 6 weeks after DOX administration (right panel), showing the preferential localization of the slow-cycling IFE cells in the bottom of epidermal undulations (stars) (**l**) and of YFP immunostaining in K14-Cre-ER/RosaYFP 8 weeks after tamoxifen administration (**m**). Dashed lines are hair follicles. Scale bars, 20 μm . Error bars show s.e.m. Hoechst nuclear staining is represented in blue.

Inv study (Fig. 2a–c), indicates that SCs are induced predominantly on cell cycle entry. This conclusion is consistent with the observation that, after administration of 12-*O*-tetradecanoyl phorbol-13-acetate (TPA),

a drug that stimulates epidermal proliferation, there is an 80% increase in clone labelling after low dose of tamoxifen (Supplementary Fig. 4a, b). Furthermore, the deceleration in the rate of clonal expansion at subsequent times indicates that, after division, SCs re-enter a quiescent phase whereas their CP cell progeny go on to proliferate and differentiate. For these clones, the quiescent SC ‘mother’ provides an anchor to the basal layer (Fig. 3a–f and Supplementary Fig. 3a), leading to clonal persistence. Such behaviour is consistent with the observation that, in K14 mice, the level of positive regulators of the cell cycle, such as *Ccnb1* and *Cdc20*, decreased, whereas negative regulators of the cell cycle, such as *Cdkn1a* and *Cdkn2a*, increased in basal cells between 3.5 days and 8 weeks (Supplementary Fig. 4c–e).

With this hypothesis, if we assume that SC division can also lead to all three possible fates (Fig. 3g), the predicted model dynamics provides an excellent fit to the clone size data (basal and suprabasal) over the wide range of time points up to 1-year post-labelling (Fig. 3h, i, Supplementary Fig. 3c, d and Supplementary Methods, section “Theory”). In particular, we find that SCs account for some $50 \pm 5\%$ of the induced proliferative cells. With a SC division rate as low as 4–6 per year, a factor of 10–20 times slower than CP cells, most of the clone size data (from 1 week to 3 months post-labelling) are largely fixed by the known characteristics of CP cells. Drawing on the early (3.5 day) time point, we are able to deduce that, in common with their progenitor cell progeny, SCs undergo population asymmetric self-renewal with some $80 \pm 10\%$ of SC divisions resulting in asymmetric fate (one SC and one CP cell), whereas the remaining divisions are equally balanced between SC duplication and symmetrical differentiation into two CPs (Fig. 3g). Finally, although the analysis points at a proliferative hierarchy, we cannot rule out the possibility that proliferative cells may switch reversibly between CP and SC behaviour if the transfer rate is sufficiently low. Moreover, the data do not allow us to infer whether the balance between SC proliferation and differentiation follows from intrinsic (cell autonomous) regulation or is controlled by environmental cues (Supplementary Methods, section “Theory”) ²¹.

Proliferation dynamics of the IFE

To challenge these findings, we applied further independent experimental assays to explore the IFE cell proliferation using a combination of cell cycle analysis, 5-bromo-2-deoxyuridine (BrdU) incorporation, and pulse-chase experiments using the K5tTA/tet(O)-H2B-GFP reporter mouse, that have proved useful in resolving quantitatively the proliferation dynamics of hair follicle bulge SCs ^{15,22–24}. We first determined the cell cycle profile of basal IFE cells by determining their DNA content. These experiments show that around 9% of $\alpha 6^+ \text{CD}34^-$ IFE cells were in the S/G2/M phase of the cell cycle (Supplementary Fig. 5a, b). Continuous BrdU administration labelled $28 \pm 2\%$ and $57 \pm 10\%$ of K5-expressing cells after 24 and 72 h, respectively, indicating an average cell division time of around 6 ± 1 days (Supplementary Fig. 5c, d and Supplementary Methods, section “Theory”), consistent with the inferred CP cell division rate.

To develop a more refined method to quantify proliferation kinetics over a longer period, we used K5tTA/tet(O)-H2B-GFP mice (Supplementary Fig. 6). In the absence of doxycycline (DOX), histone 2B domain–green fluorescent protein (H2B-GFP) fusion protein is expressed at a very high level in a discrete peak of fluorescence (Supplementary Fig. 6a–c, top panels). Following DOX administration, H2B-GFP fluorescence decreases by a factor of two after cell division ^{15,22} (Supplementary Fig. 6a–c, bottom panels). The discrete peak of H2B-GFP fluorescence observed by fluorescence-activated cell sorting (FACS) analysis during the chase period allowed an accurate quantification of the distribution in the number of rounds of division experienced by basal layer cells ($\alpha 6^+ \text{CD}34^-$) at various time points following induction (Fig. 3j and Supplementary Fig. 6d). At 3 weeks post-DOX administration, dilution of H2B-GFP label shows evidence of CP cell proliferation, while a small peak at high levels of fluorescence is consistent with around 5% or less of label-retaining cells (LRCs).

From the quantification of fluorescence over the time course, the distribution of cell divisions was shown to be quantitatively consistent with the inferred SC and CP cell dynamics (Fig. 3k and Supplementary Methods, section "Theory"). Notably, after 6 weeks of chase, the H2B-GFP label has become diluted in the vast majority of basal layer cells and only a small minority retain high levels of expression (Supplementary Fig. 6d). Intriguingly, these LRCs are localized to the bottom of the undulations of the tail epidermis (Fig. 3l), which correlate with the location of persistent clones derived by the K14 assay (Fig. 3m), and are reminiscent of the so-called rete ridges in human epidermis, where epidermal SCs were thought to reside^{25,26}.

Molecular characterization of SCs and CP cells

To gain further insight into the molecular properties of these two functionally distinct pools of epidermal progenitors, we transcriptionally profiled FACS-purified $\alpha 6^{+}CD34^{-}$ K14-Cre-ER and Inv-Cre-ER YFP-positive IFE cells 3.5 days after tamoxifen administration (Supplementary Fig. 7). Comparison of the K14 and Inv transcriptome revealed notable differences. All markers previously associated with murine and human IFE SCs^{25,27–29} were upregulated in K14-targeted cells, including $\alpha 6$, $\alpha 2$, $\alpha 3$, $\beta 1$ integrins and *Cspg4* (Supplementary Fig. 8). Real-time quantitative reverse transcriptase-PCR (qRT-PCR) performed on independent biological samples confirmed the upregulation of these integrins in K14-targeted cells (Fig. 4a). FACS analysis showed that $\alpha 6\beta 4$ integrins were expressed at a higher and more uniform level in K14- than in Inv-targeted cells. The pattern of $\alpha 2\beta 1$ integrin showed a more notable difference between the two populations, with a very high and narrow peak of $\alpha 2\beta 1$ expression in K14-targeted cells compared to the much lower and heterogeneous $\alpha 2\beta 1$ expression in Inv-targeted cells (Fig. 4b, c and Supplementary Fig. 9). Gene ontology classification of the genes differentially expressed by more than 1.5-fold between K14- and Inv-labelled cells revealed that genes preferentially expressed in

K14-positive cells are enriched with a very high statistical significance in functional groups involved in positive and negative cell cycle regulation (for example, *Ccn2*, *Cdc20*, *Cdkn1a*, *Cdkn2a*), mitosis (for example, *Ccnb1*, *Kif11*), chromosome segregation (for example, *Cenpe*), regulation of cell proliferation (for example, *Epgn*), and DNA repair (for example, *Brca1*) (Fig. 4a, Supplementary Figs 8, 10a and Supplementary Table 1). By contrast, Inv-Cre-ER labelled cells preferentially expressed genes known to control epidermal differentiation^{30–32} (for example, *Notch3*, *Grhl3*), keratinocyte differentiation and keratinization (for example, *Sppr1a*, *1b*, *2d*, *2i*)³³, and lipid metabolism (for example, *Elovl4*, *6*, *7*, *Olah*), as confirmed by qRT-PCR analysis (Fig. 4d, Supplementary Figs 8, 10b and Supplementary Table 1). Altogether these data demonstrate that K14- and Inv-targeted cells differ by their gene expression profile, and uncover many new markers preferentially expressed by IFE SC and CP cells.

To characterize further the molecular heterogeneity of basal IFE cells, we FACS-purified different populations of basal epidermal cells ($\alpha 6^{+}CD34^{-}$) according to the level of $\beta 1$ integrin expression which, together with $\alpha 2$ integrin, present the most significant difference between K14 and Inv-targeted cells at 3.5 days following tamoxifen administration (Fig. 4b and Supplementary Fig. 9). We fractionated epidermis into three distinct populations: one population (<3%) expressing very high levels of $\beta 1$ integrin (which are likely to be enriched in SCs as they are only detected in K14- but not Inv-targeted cells), a second population expressing intermediate levels of $\beta 1$ integrin detected in both K14 and Inv-targeted cells (which are likely to represent CP cells), and cells expressing low levels of $\beta 1$ integrin only found in Inv-targeted cells (which are likely to represent the differentiated basal cells and early suprabasal cells) (Supplementary Fig. 11a). Strikingly, many markers identified as being upregulated in K14 versus Inv were also upregulated in cells expressing very high levels of $\beta 1$ integrin compared to cells expressing intermediate or low levels (Supplementary Fig. 11b). Similarly, many genes preferentially expressed by Inv cells were already upregulated in cells expressing intermediate levels of $\beta 1$ integrin compared to $\beta 1$ integrin high cells but, nevertheless, much less expressed than in cells expressing low levels of $\beta 1$ integrin (Supplementary Fig. 11c). These data demonstrate that the expression of many differentiation-associated genes is already upregulated at the earliest stage of progenitor commitment, reminiscent of the lineage priming reported in haematopoietic stem cells^{34,35}.

Long-term SC contribution to wound healing

In addition to their role in maintenance, adult SCs are implicated in the repair of damaged tissue following injury. Lineage tracing has shown that hair follicle SCs are mobilized to the wound area and contribute to the repair of damaged epidermis^{15–18}. At the same time, genetic mouse mutants, that present a complete absence of hair follicle, heal incisional wounds with a slight delay of the reepithelialisation, demonstrating that IFE cells are also capable of tissue regeneration³⁶. However, in unperturbed conditions, the contribution of IFE SCs and progenitors to wound healing is currently unknown. To address the respective contributions of K14-SCs and Inv-CPs during tissue repair, we first marked basal progenitors by titrating the dose of tamoxifen so as to label Inv and K14-Cre-ER epidermis at roughly the same density (4.5 ± 0.1 clones per mm^2 and 3.8 ± 0.5 clones per mm^2 in K14 and Inv, respectively), then subjected tail epidermis to full thickness excisional wound using punch biopsy, and assessed the contribution of marked cells to repair (Fig. 5a). Whole-mount analysis showed major recruitment of K14-Cre-ER-labelled IFE cells to the wound area, with clones migrating from the periphery towards the centre of the wound (with only 7.5% of clones from hair follicle), persisting long-term after wounding (2.6 ± 0.1 clones per mm^2 after 35 days) (Fig. 5b, c and Supplementary Fig. 12), and consisting of very large clones with a broad basal attachment and huge numbers of differentiated cells (Fig. 5d). In sharp contrast, few Inv-Cre-ER-targeted cells were recruited to the wound area, and these clones were, on average, much

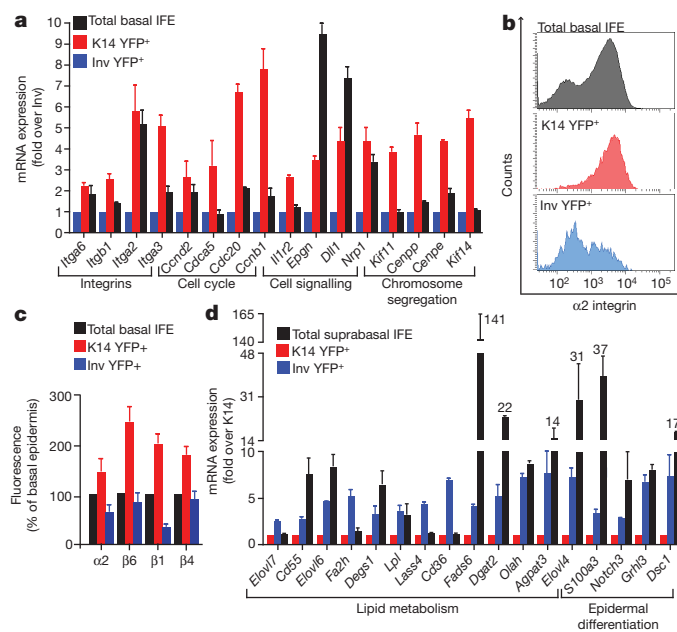


Figure 4 | Molecular signature of K14 SC and Inv CP cells. **a**, qRT-PCR analysis of the expression of representative genes upregulated in K14-Cre-ER-targeted cells, compared with the level of expression of the same genes in Inv-Cre-ER-targeted cells and total basal IFE cells ($n = 3$). **b**, **c**, Fluorescence histogram (**b**) and geometric mean fluorescence (**c**) of integrin cell-surface expression in all basal IFE cells, K14-Cre-ER- and Inv-Cre-ER-targeted IFE cells, as determined by FACS analysis, 3.5 days after tamoxifen ($n = 3$). **d**, qRT-PCR analysis of the expression of representative genes upregulated in Inv-Cre-ER-targeted cells, compared to the level of expression of the same genes in basal epidermal K14-Cre-ER-targeted cells and total suprabasal IFE cells ($n = 3$). Error bars show s.e.m.

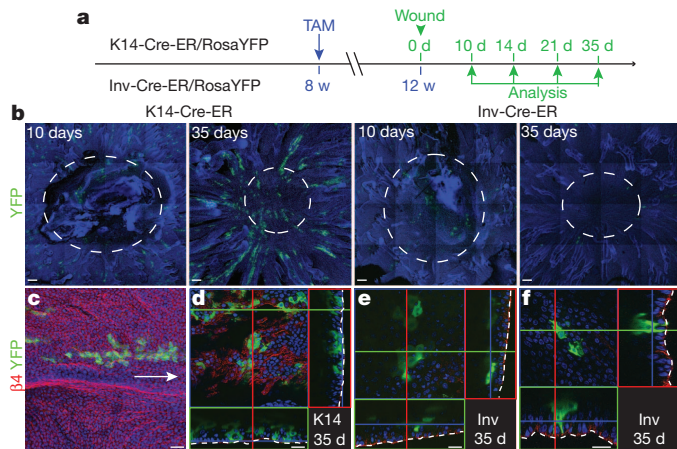


Figure 5 | Massive and sustained contribution of K14 SC during wound healing. **a**, Scheme representing the experimental strategy used to label basal progenitors and assess their contribution during wound healing. **b**, Microscopic analysis of YFP immunostaining performed on whole-mount of wounded tail epidermis 4 weeks after tamoxifen administration to K14-Cre-ER/RosaYFP and Inv-Cre-ER/RosaYFP mice and analysed at different times following wounding. Dashed lines represent the wounded area. Scale bar, 200 μ m. **c**, Confocal analysis of whole-mount immunostaining of YFP and β 4 integrin 21 days following wounding, showing the departure of K14 YFP labelled from the IFE. The arrow indicates the wound localization. Scale bars, 20 μ m. **d–f**, Confocal analysis of representative clones from wounded tail epidermis 35 days post-wound, showing very big clones containing multiple basal cells in K14 targeted cells (**d**), whereas the rare still visible Inv marked clones are either completely detached from the basal epidermis (**e**) or very small containing one or few basal cells (**f**). Dashed lines represent the basal lamina. Scale bars, 20 μ m. Hoechst nuclear staining is represented in blue.

smaller. At 35 days post-injury, few Inv clones survived (0.3 ± 0.3 clones per mm^2) and consisted of either clones of cells being shed with no basal attachment (Fig. 5e), or small clones with one or few basal cells (Fig. 5f). These data demonstrate that K14-derived SC are capable of extensive tissue regeneration, whereas Inv CP cells have only a limited contribution to wound healing.

Discussion

Our study demonstrates the existence, hierarchical organization and the proliferation dynamics of two distinct types of progenitors that effect different functions during homeostasis and repair of the IFE in adult mice. Quantitative modelling of clonal fate data suggest that both the slow-cycling SCs and the more rapidly cycling CP cells share a similar pattern of population asymmetric self-renewal in which the balance between proliferation and differentiation is achieved through stochastic fate choice (Supplementary Fig. 13). These findings provide a reconciliation of seemingly contradictory theories of IFE maintenance^{4,12}, and explain the proliferative heterogeneity previously reported in IFE^{5,37,38}. The existence of slow-cycling SCs and more rapidly cycling progenitors is reminiscent of the situation encountered in other tissues including cornea³⁹, hair follicle⁴⁰, blood^{41,42}, muscle⁴³ and brain⁴⁴. In all of these tissues, these slow-cycling SCs can switch rapidly and reversibly between quiescence and activity following injury and/or drug treatment^{15,45}. Indeed, this partitioning of function—progenitors to undertake routine maintenance and quiescent SCs to effect repair—may represent a generic strategy of tissue maintenance.

The demonstration that the long-term wound healing potential is dominated by the IFE SC compartment may suggest that either SC and CP cell populations are not interchangeable, in contrast to progenitors in intestinal epithelium⁴⁶, or that their reversion rate is sufficiently slow to render its effects negligible. The seemingly short-lived contribution of Inv CP cells to the repair of the epidermis is reminiscent of the fate of K15 bulge-derived SCs during wounding¹⁶, and

demonstrates the critical role of IFE SCs for the long-term repair of the epidermis, as has been proposed¹⁶.

METHODS SUMMARY

Clonal YFP expression in the tail IFE was performed by administering 0.2 mg and 2.5 mg tamoxifen to K14-Cre-ER/RosaYFP and Inv-Cre-ER/RosaYFP mice, respectively. Immunostainings were performed as described⁴⁷. Quantification of the proportion of surviving clones as well as the basal, suprabasal and total clone size was determined by counting the number of YFP⁺ cells using whole-mount tail epidermis, analysed by confocal microscopy. BrdU pulse and H2B-GFP label retention was performed as described^{15,43}. The wound healing assay was performed by administering 1 mg and 10 mg tamoxifen to K14-Cre-ER/RosaYFP and Inv-Cre-ER/RosaYFP mice, respectively. Four weeks later punch biopsies were performed in the tail epidermis and analysed at different time points. Mathematical modelling of the clonal fate data was performed as described in Supplementary Methods, section 'Theory'. For further details see Supplementary Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 13 March; accepted 3 July 2012.

Published online 2 September 2012.

- Blanpain, C. & Fuchs, E. Epidermal homeostasis: a balancing act of stem cells in the skin. *Nature Rev. Mol. Cell Biol.* **10**, 207–217 (2009).
- Potten, C. S., Saffhill, R. & Maibach, H. I. Measurement of the transit time for cells through the epidermis and stratum corneum of the mouse and guinea-pig. *Cell Tissue Kinet.* **20**, 461–472 (1987).
- Potten, C. S. Cell replacement in epidermis (keratopoiesis) via discrete units of proliferation. *Int. Rev. Cytol.* **69**, 271–318 (1981).
- Potten, C. S., Wichmann, H. E., Loeffler, M., Dobek, K. & Major, D. Evidence for discrete cell kinetic subpopulations in mouse epidermis based on mathematical analysis. *Cell Tissue Kinet.* **15**, 305–329 (1982).
- Potten, C. S. & Loeffler, M. Epidermal cell proliferation. I. Changes with time in the proportion of isolated, paired and clustered labelled cells in sheets of murine epidermis. *Virchows Arch. B Cell Pathol. Incl. Mol. Pathol.* **53**, 279–285 (1987).
- Mackenzie, I. C. Retroviral transduction of murine epidermal stem cells demonstrates clonal units of epidermal structure. *J. Invest. Dermatol.* **109**, 377–383 (1997).
- Kolodka, T. M., Garlick, J. A. & Taichman, L. B. Evidence for keratinocyte stem cells *in vitro*: long term engraftment and persistence of transgene expression from retrovirus-transduced keratinocytes. *Proc. Natl Acad. Sci. USA* **95**, 4356–4361 (1998).
- Ghazizadeh, S. & Taichman, L. B. Multiple classes of stem cells in cutaneous epithelium: a lineage analysis of adult mouse skin. *EMBO J.* **20**, 1215–1222 (2001).
- Ghazizadeh, S. & Taichman, L. B. Organization of stem cells and their progeny in human epidermis. *J. Invest. Dermatol.* **124**, 367–372 (2005).
- Ro, S. & Rannala, B. A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep.* **5**, 914–920 (2004).
- Ro, S. & Rannala, B. Evidence from the stop-EGFP mouse supports a niche-sharing model of epidermal proliferative units. *Exp. Dermatol.* **14**, 838–843 (2005).
- Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446**, 185–189 (2007).
- Doupé, D. P., Klein, A. M., Simons, B. D. & Jones, P. H. The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. *Dev. Cell* **18**, 317–323 (2010).
- Jones, P. & Simons, B. D. Epidermal homeostasis: do committed progenitors work while stem cells sleep? *Nature Rev. Mol. Cell Biol.* **9**, 82–88 (2008).
- Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
- Ito, M. *et al.* Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. *Nature Med.* **11**, 1351–1354 (2005).
- Levy, V., Lindon, C., Zheng, Y., Harfe, B. D. & Morgan, B. A. Epidermal stem cells arise from the hair follicle after wounding. *FASEB J.* **21**, 1358–1366 (2007).
- Snippert, H. J. *et al.* Lgr6 marks stem cells in the hair follicle that generate all cell lineages of the skin. *Science* **327**, 1385–1389 (2010).
- Vasioukhin, V., Degenstein, L., Wise, B. & Fuchs, E. The magical touch: genome targeting in epidermal stem cells induced by tamoxifen application to mouse skin. *Proc. Natl Acad. Sci. USA* **96**, 8551–8556 (1999).
- Lapouge, G. *et al.* Identifying the cellular origin of squamous skin tumors. *Proc. Natl Acad. Sci. USA* **108**, 7431–7436 (2011).
- Klein, A. M. & Simons, B. D. Universal patterns of stem cell fate in cycling adult tissues. *Development* **138**, 3103–3111 (2011).
- Waghmare, S. K. *et al.* Quantitative proliferation dynamics and random chromosome segregation of hair follicle stem cells. *EMBO J.* **27**, 1309–1320 (2008).
- Zhang, Y. V., Cheong, J., Ciapurin, N., McDermitt, D. J. & Tumbar, T. Distinct self-renewal and differentiation phases in the niche of infrequently dividing hair follicle stem cells. *Cell Stem Cell* **5**, 267–278 (2009).
- Zhang, Y. V., White, B. S., Shalloway, D. I. & Tumbar, T. Stem cell dynamics in mouse hair follicles: A story from cell division counting and single cell lineage tracing. *Cell Cycle* **9**, 1504–1510 (2010).

25. Jones, P. H., Harper, S. & Watt, F. M. Stem cell patterning and fate in human epidermis. *Cell* **80**, 83–93 (1995).
26. Lavker, R. M. & Sun, T. T. Heterogeneity in epidermal basal keratinocytes: morphological and functional correlations. *Science* **215**, 1239–1241 (1982).
27. Jones, P. H. & Watt, F. M. Separation of human epidermal stem cells from transit amplifying cells on the basis of differences in integrin function and expression. *Cell* **73**, 713–724 (1993).
28. Legg, J., Jensen, U. B., Broad, S., Leigh, I. & Watt, F. M. Role of melanoma chondroitin sulphate proteoglycan in patterning stem cells in human interfollicular epidermis. *Development* **130**, 6049–6063 (2003).
29. Tani, H., Morris, R. J. & Kaur, P. Enrichment for murine keratinocyte stem cells based on cell surface phenotype. *Proc. Natl Acad. Sci. USA* **97**, 10960–10965 (2000).
30. Rangarajan, A. *et al.* Notch signaling is a direct determinant of keratinocyte growth arrest and entry into differentiation. *EMBO J.* **20**, 3427–3436 (2001).
31. Blanpain, C., Lowry, W. E., Pasolli, H. A. & Fuchs, E. Canonical notch signaling functions as a commitment switch in the epidermal lineage. *Genes Dev.* **20**, 3022–3035 (2006).
32. Ting, S. B. *et al.* A homolog of *Drosophila* grainy head is essential for epidermal integrity in mice. *Science* **308**, 411–413 (2005).
33. Candi, E., Schmidt, R. & Melino, G. The cornified envelope: a model of cell death in the skin. *Nature Rev. Mol. Cell Biol.* **6**, 328–340 (2005).
34. Månsson, R. *et al.* Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**, 407–419 (2007).
35. Pina, C. *et al.* Inferring rules of lineage commitment in haematopoiesis. *Nature Cell Biol.* **14**, 287–294 (2012).
36. Langton, A. K., Herrick, S. E. & Headon, D. J. An extended epidermal response heals cutaneous wounds in the absence of a hair follicle stem cell contribution. *J. Invest. Dermatol.* **128**, 1311–1318 (2008).
37. Loeffler, M., Potten, C. S. & Wichmann, H. E. Epidermal cell proliferation. II. A comprehensive mathematical model of cell proliferation and migration in the basal layer predicts some unusual properties of epidermal stem cells. *Virchows Arch. B Cell Pathol. Incl. Mol. Pathol.* **53**, 286–300 (1987).
38. Morris, R. J., Fischer, S. M. & Slaga, T. J. Evidence that the centrally and peripherally located cells in the murine epidermal proliferative unit are two distinct cell populations. *J. Invest. Dermatol.* **84**, 277–281 (1985).
39. Cotsarelis, G., Cheng, S. Z., Dong, G., Sun, T. T. & Lavker, R. M. Existence of slow-cycling limbal epithelial basal cells that can be preferentially stimulated to proliferate: implications on epithelial stem cells. *Cell* **57**, 201–209 (1989).
40. Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61**, 1329–1337 (1990).
41. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
42. Foudi, A. *et al.* Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nature Biotechnol.* **27**, 84–90 (2009).
43. Rocheteau, P., Gayraud-Morel, B., Siegl-Cachedenier, I., Blasco, M. A. & Tajbakhsh, S. A subpopulation of adult skeletal muscle stem cells retains all template DNA strands after cell division. *Cell* **148**, 112–125 (2012).
44. Bonaguidi, M. A. *et al.* *In vivo* clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics. *Cell* **145**, 1142–1155 (2011).
45. Essers, M. A. *et al.* IFN α activates dormant haematopoietic stem cells *in vivo*. *Nature* **458**, 904–908 (2009).
46. Takeda, N. *et al.* Interconversion between intestinal stem cell populations in distinct niches. *Science* **334**, 1420–1424 (2011).
47. Youssef, K. K. *et al.* Identification of the cell lineage at the origin of basal cell carcinoma. *Nature Cell Biol.* **12**, 299–305 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank F. Bollet-Quivogne and J.-M. Vanderwinden for their help with confocal imaging. C.B. and P.A.S. are chercheur qualifié, G.M. and S.B. are supported by fellowship of the FRS/FNRS. B.D. is supported by TELEVIE. C.B. is an investigator of WELBIO. This work was supported by the FNRS, the program d'excellence CIBLES of the Wallonia Region, a research grant from the Fondation Contre le Cancer, the ULB foundation, the fond Gaston Ithier, the European Research Council (ERC) and the EMBO Young Investigator Program.

Author Contributions C.B., G.M., B.D., S.D., P.A.S. and B.D.S. designed the experiments and performed data analysis. G.M., S.D., B.D. and K.K.Y. performed all the experiments. S.B. performed bioinformatic analysis of the microarray. C.B. and B.D.S. wrote the manuscript.

Author Information The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus and are accessible through GEO Series accession number GSE36688. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.B. (cedric.blanpain@ulb.ac.be) or B.D.S. (bds10@cam.ac.uk).

METHODS

Mice. K14-Cre-ER¹⁹ and TRE-mCMV-H2B-GFP¹⁵ transgenic mice were provided by E. Fuchs. K5-tTA mice⁴⁸ were a gift from A. Glick. RosaYFP mice⁴⁹ were obtained from Jackson Laboratory. Involucrin-CreERT2 were previously described²⁰. Mouse colonies were maintained in a certified animal facility in accordance with European guidelines. These experiments were approved by the local ethical committee (CEBEA).

Targeting YFP expression. For lineage tracing experiment, K14-Cre-ER/RosaYFP and Involucrin-CreERT2/Rosa-YFP mice were induced between 2 and 4 months with 0.2 mg and 2.5 mg of tamoxifen (Sigma-Aldrich), respectively, by intraperitoneal injection.

For cell sorting, K14-Cre-ER/RosaYFP and Inv-Cre-ER/RosaYFP mice were induced with 1 mg and 10 mg of tamoxifen, respectively, by intraperitoneal injection and isolated 4 days later.

Proliferation experiments. For H2B-GFP chase, K5-Tet off/TRE-mCMV-H2B-GFP were treated once with 200 μ l of doxycycline (2 mg ml⁻¹) by intraperitoneal injection and simultaneously continually fed with doxycycline food (1 g kg⁻¹) until animal euthanasia. For BrdU experiment, mice were injected every 12 h with 200 μ l of BrdU (10 mg ml⁻¹) and analysed 24 h and 72 h after the first injection.

Wound experiments. After mice anaesthesia (5% xylazine 10% ketamine in PBS), circular pieces of epidermis were removed from K14-Cre-ER/Rosa YFP and Inv-Cre-ER/RosaYFP tail epidermis, 4 weeks after tamoxifen induction, using a 3 mm diameter biopsy punch (Stiefel).

Histology and immunostaining. Skin epidermis was removed from tail bone and pre-fixed overnight in 4% paraformaldehyde at 4 °C. Tissues were washed three times in PBS for 5 min and incubated overnight in 30% sucrose in PBS at 4 °C. Tissues were then embedded in OCT and kept at -80 °C. Sections of 6 μ m were cut using a CM3050S Leica cryostat (Leica Microsystems).

Sections were incubated in blocking buffer (1% BSA, 5% horse serum, 0.2% Triton in PBS) for 1 h at room temperature. Primary antibodies were incubated overnight at 4 °C or 1 h at room temperature in the dark. Sections were rinsed three times in PBS and incubated with appropriate secondary antibodies diluted to 1:400 and Hoeschst in blocking buffer for 1 h at room temperature. Sections were again washed three times with PBS. The following primary antibodies were used: anti-involucrin (rabbit, 1:1,000, Covance), anti-K5 (rabbit, 1:1,000, Covance), anti-K10 (rabbit, 1:1,000, Covance), anti- β 4 (rat, 1:200, BD Biosciences), anti-GFP (rabbit, 1:1,000, Molecular Probes), anti-GFP (chicken, 1:1,000, Abcam). The following secondary antibodies were used: anti-rabbit, anti-rat, anti-goat, anti-chicken conjugated to Alexa Fluor 488 (Molecular Probes), to rhodamine Red-X (Jackson ImmunoResearch) or to Cy5 (Jackson ImmunoResearch). Nuclei were stained in Hoechst solution (1:2,000) and slides were mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

Epidermal whole-mount. Pieces of skin tail were incubated in EDTA (20 mM) on a rocking plate at 37 °C for 1 h. Epidermis was separated from the dermis as an intact sheet and washed 2 times with PBS. Pieces of epidermis were pre-fixed in 4% paraformaldehyde overnight at 4 °C or 1 h at room temperature. Epidermis were rinsed 2 times with PBS for 5 min and conserved in PBS with 0.2% azide at 4 °C. Small pieces of epidermis were incubated in blocking buffer (1% BSA, 5% horse serum, 0.8% Triton in PBS) for 3 h at room temperature on a rocking plate (100 r.p.m.). The samples were incubated in primary antibodies overnight at room temperature on the rocking plate. The primary antibodies used were the following: anti-K5 (rabbit, 1:200, Covance), anti- β 4 (rat, 1:100, BD Biosciences), anti-GFP (rabbit, 1:200, Molecular Probes), anti-GFP (chicken, 1:200, Abcam). Samples were then washed 3 times in PBS with 0.2% tween for 1 h and incubated in appropriate secondary antibodies diluted 1:400 in blocking buffer overnight at 4 °C on the rocking plate. Pieces of epidermis were then washed with PBS 3 times for 1 h. For BrdU staining, samples were incubated in HCl 1 M at 37 °C for 20 min, washed with PBS 0.2% tween, stained with anti-BrdU (rat, 1:200, Abcam) in blocking buffer and with appropriate secondary antibody after several washes. Nuclei were stained in Hoechst solution diluted 1:1,000 for 30 min and mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

TPA experiment. 12-O-tetradecanoyl-phorbol-13-acetate (TPA) from Sigma has been diluted in acetone (final concentration 25 mg ml⁻¹). K14-Cre-ER/RosaYFP mice were topically treated with TPA on tail epidermis during 11 days, induced with tamoxifen the 8th days. The 12th day, treated mice were killed and processed to get whole-mounts of tail epidermis.

Microscope image acquisition and quantification. All pictures of section immunostaining were acquired using the Axio Imager M1 Microscope, the AxioCamMR3 or MrC5 camera and using the Axiovision software (Carl Zeiss). Acquisitions were performed at room temperature using $\times 20$ numerical aperture (NA) 0.4 and $\times 40$ NA 0.75 EC Plan-Neofluar objectives (Carl Zeiss). All confocal pictures on whole-mount were acquired at room temperature using a Zeiss

LSM780 confocal microscope fitted on an Axiovert M200 inverted microscope equipped with a $\times 40$ NA 1.2 C-Apochromat water immersion objective (Carl Zeiss MicroImaging GmbH). Sequential scanings of 1,024 \times 1,024 pixels, z-stack 0.3 μ m, were acquired using the ZEN 2010 software (Carl Zeiss).

Number of basal clones per interfollicular region were quantified on frozen section using Axio Imager M1 microscope for each time point and represented as percentage of basal clone. Clone sizes were quantified on whole-mount tail epidermis by acquiring individual clones using LSM 780 confocal microscope and the number of basal and suprabasal cells were counted.

Dissociation of epidermal cells. Skin epidermis (from CD1, K5tTA/tet(o)-H2BGFP treated or not with DOX, K14-Cre-ER and Inv-Cre-ER/RosaYFP induced with tamoxifen for 3.5 days or 8 weeks) was removed from tail bone and incubated overnight in HBSS (Gibco) 0.25% trypsin (Gibco) at 4 °C. Epidermis was separated from the dermis and incubated on a rocking plate (100 r.p.m.) at room temperature for 5 min. Basal cells were mechanically separated from the epidermis by flushing 10 times under the epidermis. Tissues were then cut in pieces of 1 mm² with scalpel and trypsin was neutralized by adding DMEM medium (Gibco) supplemented with 2% Chelex fetal calf serum (FCS). Samples were filtrated on 70- and 40- μ m filter (Falcon).

Cell labelling, flow cytometry and cell sorting. Using a concentration of 20 millions cells per ml, cells were incubated in 2% FCS/PBS with primary antibodies for 30 min on ice, protected from the light, with shaking every 10 min. Immunostaining for isolating K14 and involucrin cells was performed using biotin-conjugated anti-CD34 (clone RAM34; BD Biosciences) and phycoerythrin-conjugated anti- α 6 integrin (clone GoH3; BD Biosciences). Primary antibodies were washed with 2% FCS/PBS and cells incubated for 30 min in allophycocyanin-conjugated streptavidin (BD Biosciences) secondary antibodies, on ice, with shaking every 10 min. Living K14 and involucrin expressing epidermal cells were gated by forward scatter, side scatter, negative staining for Hoechst dye and by following the YFP signal. Basal cells from the interfollicular epidermis were targeted using CD34-negative α 6-integrin-positive gating. Immunostaining for isolating basal β 1 high, intermediate and low cells was performed using biotin-conjugated anti-CD34 revealed with a phycoerythrin-Cy7-conjugated streptavidin (BD Biosciences) and anti-CD29 allophycocyanin-conjugated (β 1 integrin, clone Hmb1-1, eBiosciences). Single living, CD34-negative, β 1 low/medium/high cells were sorted. β 1 low, medium and high gates were set up on the basis of its expression in K14-Cre-ER- and Inv-Cre-ER/RosaYFP-targeted basal IFE cells. Fluorescence-activated cell sorting analysis was performed using FACSAria I at high pressure (70 p.s.i.) and FACSDiva software (BD Biosciences). Sorted cells were harvested directly in the lysis buffer provided by the RNeasy microkit (QIAGEN) supplemented with 1 μ l of beta-mercaptoethanol for every 100 μ l of lysis buffer. RNA extraction was performed on freshly sorted cells according to the manufacturer's protocol. The entire procedure was repeated in at least three biologically independent samples. Other integrin antibodies were used to determine their expression in the different cell populations: anti-CD49b phycoerythrin-conjugated (α 2 integrin, clone Hm α 2, BD Biosciences), anti-CD29 allophycocyanin-conjugated, and anti-CD104 rat (β 4 integrin, clone 346 11A, BD Biosciences) revealed with an anti-rat Rhodamine Red x-conjugated (donkey, JacksonImmunoResearch).

Total suprabasal and total basal IFE cells isolation for qPCR controls. For total suprabasal IFE control, skin epidermis was removed from tail bone and incubated 1 h in EDTA (20 mM) on a rocking plate at 37 °C. Epidermis was separated from the dermis and basal cells were flushed out from the epidermis. The piece of epidermis containing the suprabasal cells was then frozen in liquid nitrogen. RNA extraction was performed using RNeasy microkit (QIAGEN): the tissue was cut in small pieces and lysed in RLT buffer containing β -mercaptoethanol according to the manufacturer's protocol. For total basal IFE qPCR control, cells were obtained by gating the whole α 6⁺ (low medium and high) and CD34⁻ population in FACS sorting.

RNA quality. Extracted RNA quality was tested using capillary electrophoresis (Agilent Bioanalyzer, Agilent RNA 6000 Nano Kit). Degraded RNA samples or presenting sign of degradation or containing high proteins or salt contamination were discarded.

Microarray analysis. Cohort of 10 Inv-Cre-ER/RosaYFP and K14-Cre-ER/RosaYFP mice were used to isolate at least 100,000 basal IFE cells (α 6⁺ CD34⁻ YFP⁺) per replicate. Total RNAs were isolated from these sorted cells and were labelled and hybridized on a mouse genome 430 2.0 array. Microarrays were performed in duplicate for the untreated K14-Cre-ER/RosaYFP and tamoxifen-treated K14-Cre-ER/RosaYFP and Inv-Cre-ER/RosaYFP sorted cells. All the results were normalized using the frozen robust multiarray analysis (fRMA) normalization using R-bioconductor package fRMA^{50,51} with standard parameters. Genetic signatures were obtained by considering genes presenting a fold change greater on smaller than 2 or -2, respectively.

Reverse transcription and quantitative PCR. Cohort of 10 Inv-Cre-ER/RosaYFP and K14-Cre-ER/RosaYFP mice were used to isolate at least 100,000 basal IFE cells ($\alpha 6^{+}$ CD34 $^{-}$ YFP $^{+}$) per replicate. Each RNA sample was quantified using a nanodrop spectrophotometer. RNA quality and quantity used for qRT-PCR was exactly the same as for the RNA used for Microarray analysis. Purified RNA (200 ng) was used to synthesize the first-strand cDNA using Superscript II (Invitrogen) and random hexamers (Roche). Mock was obtained following the same procedure without adding Superscript II. Quantitative PCR analyses were performed with 1 ng of complementary DNA as template, using a Brilliant II Fast SYBR QPCR Master Green mix (Stratagene) and an Agilent Technologies Stratagene Mx3500P real-time PCR system.

All primers were designed using the Assay Design Center (Roche applied science, https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp?id=uplct_030000). The list of primers are shown in the table underneath. The linearity of each pair of primer used has been tested with the following dilution of cDNA: 8 ng, 2 ng, 0.5 ng, 0.125 ng.

Analysis of results was performed with Mxpro software (Stratagene). Delta delta CT were used to calculate the relative expression of involucrin samples to the K14 samples using the housekeeping gene TataBox.

Primers. Gene symbol, forward primer and reverse primer (5' to 3') are: *Agpat3*, ccacagcaagcacctatacc and cactccaggagcatgacca; *Ccnb1*, tgcattttgctccttctcaa and caggaagcaggaggtcttca; *Ccnd2*, caccgacaactctgtgaagc and tcacttcagcttaccaca; *Cd36*, ttgtacctatactgtggctaaatgaga and cttgtgttttgaacatttctgctt; *Cd55*, actgttgattggg acgatgag and tgggtgctctggacaatgta; *Cdc20*, acatcaaggcgtgtcaag and aatgtgccggtc actggt; *Cdca5*, cacagtgtatgttaggaactggc and caccctgtcaccaatgt; *Cenpe*, tctttaccgtctgaggtggaa and ggagctcttcagatttctcatata; *Cenpp*, gcgaatgctgcaaaacg and caaagattcccactcctcaga; *Degs1*, tcacttcaatgtgggctatc and cacttgcatcttctcacc; *Dgat2*, ggcgctacttccgagactac and ttgtcagcaggtgtgtgtgt; *Dll1*, gggcttctctgcttcaac

and taagagttgccgaggtccac; *Dsc1*, gggagcacttctctaaagca and ttttgacaggcatcacaataa; *Elovl4*, acgacaccgtggagttctatc and gcggccagctgtctacac; *Elovl6*, cagcaagcaccgc aacta and aggagcacagtgtgtgggtg; *Elovl7*, tctcagtcgccaagagcaa and acagctcgatga atttgaga; *Epgn*, ggctctgggggttctgatag and cctctgcttctcgtcag; *Fa2h*, tgggtgactggg acaagg and gttgtgtgaaccactcatca; *Fads6*, ccaccttatctccatgtcaa and cttgtccggggag aacatag; *Grhl3*, aaggaagatgtcgaatgaacttg and tcgtcctcattactgtaggga; *Il1r2*, cccatccctgtgatcatttc and gcacgggactatcagtttga; involucrin, atgtcccatcaacacacactg and atgtcccatcaacacacactg; *Itga2*, gggaccggaggcctttcta and tgcataatactgatttccacactg; *Itga3*, aggatattgtgcttgagtgga and gaccacagcacttgggtga; *Itga6*, attcaggagtagcttggtg gat and cttatttctctgaagaagccacac; *Itgb1*, atgcaggttgcggttgt and catcgtggaaaacacc ag; *Kif11*, aaagaggaaaaggcaggaa and cgcctctgccagattaatg; *Kif14*, ggctatttgcaat tgttttgc and ttgagcctctttaaccatcg; *Kif2c*, cgaaggaggtaccacaaaagg and ttcggtc gtaagggaagaag; *Lass4*, gcctgcatcttcttcttg and ctgccacagccactcactc; *Lpl*, ctggt gggaaatgatgtgg and tggacgttctctagggggta; *Notch3*, agctgggtcctgaggtgat and agacag agccggtgtcaat; *Nrp1*, gaggaatgttctgtcgtatga and ccaatgtgagggccaact; *Olah*, ggccgaaagattaatggctta and ggttcttcaaatcgggttct; *S100a3*, gggacaccagttggttagg and gcacacgatggcagctact; T-box, gtaccgcagcttcaaaatattgtat and aaatcaacgcag ttgtcgtg.

48. Diamond, I., Owolabi, T., Marco, M., Lam, C. & Glick, A. Conditional gene expression in the epidermis of transgenic mice using the tetracycline-regulated transactivators tTA and rTA linked to the keratin 5 promoter. *J. Invest. Dermatol.* **115**, 788–794 (2000).
49. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1**, 4 (2001).
50. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
51. McCall, M. N., Bolstad, B. M. & Irizarry, R. A. Frozen robust multiarray analysis (fRMA). *Biostatistics* **11**, 242–253 (2010).

Ploughing the deep sea floor

Pere Puig¹, Miquel Canals², Joan B. Company¹, Jacobo Martín¹, David Amblas², Galderic Lastras², Albert Palanques¹ & Antoni M. Calafat²

Bottom trawling is a non-selective commercial fishing technique whereby heavy nets and gear are pulled along the sea floor. The direct impact of this technique on fish populations^{1,2} and benthic communities^{3,4} has received much attention, but trawling can also modify the physical properties of seafloor sediments, water–sediment chemical exchanges and sediment fluxes^{5,6}. Most of the studies addressing the physical disturbances of trawl gear on the seabed have been undertaken in coastal and shelf environments^{7,8}, however, where the capacity of trawling to modify the seafloor morphology coexists with high-energy natural processes driving sediment erosion, transport and deposition⁹. Here we show that on upper continental slopes, the reworking of the deep sea floor by trawling gradually modifies the shape of the submarine landscape over large spatial scales. We found that trawling-induced sediment displacement and removal from fishing grounds causes the morphology of the deep sea floor to become smoother over time, reducing its original complexity as shown by high-resolution seafloor relief maps. Our results suggest that in recent decades, following the industrialization of fishing fleets, bottom trawling has become an important driver of deep seascape evolution. Given the global dimension of this type of fishery, we anticipate that the morphology of the upper continental slope in many parts of the world's oceans could be altered by intensive bottom trawling, producing comparable effects on the deep sea floor to those generated by agricultural ploughing on land.

Continental slopes are at the transition between shallow continental shelves and deep basins and are often cut by submarine canyons. They represent the steepest part of continental margins, which extend from the shelf break, usually at depths of 100–200 m, to the upper continental rise at depths of 1,500–3,000 m. At fine spatial and temporal scales (that is, excluding tectonics), continental-slope morphological features form as a result of mechanisms governed by

erosional and depositional processes¹⁰. Slope failures and sediment gravity flows¹¹, geostrophic contour-flowing bottom currents¹², internal waves¹³ and dense shelf water cascading¹⁴ are some of the main processes shaping continental slopes.

Bottom trawling is recurrently practised in numerous continental slopes of the world's oceans (Fig. 1). Driven by technical improvements and the exhaustion of traditional fish stocks, trawling fleets have progressively expanded their activities from the continental shelf towards deeper regions^{15,16}, where background energy levels controlling natural sedimentary processes are generally low except for infrequent, localized events^{11,14}.

The northwestern Mediterranean continental slope is a region where deep-sea trawling has been practised for a long time. A monospecific fishery targeting blue and red deep-sea shrimp *Aristeus antennatus* (Risso, 1816) was active in this area as early as the 1930s (ref. 17). However, fishing effort (a measure of the amount of fishing, generally defined as the fleet capacity times days at sea) was negligible until the mid-1960s, when the trawling fleet became fully industrialized. This trend notably accelerated in the 1970s in terms of engine power¹⁸ and, subsequently, of gear size, fishing depth and trawled area per haul. The slope fishing grounds in the northern Catalan margin are currently being regularly exploited by 180 large bottom trawlers (that is, each over 15 m in overall length) down to 800 m depth, with greater depths being reached occasionally (Fig. 2).

Time series observations were carried out on the flank of a trawled submarine canyon (La Fonera Canyon, Fig. 2) to measure sediment transport induced by fishing activities. Following an earlier study on sedimentary processes¹⁹, an instrumented mooring was placed at 980 m depth within a tributary valley on the canyon flank to capture the occurrence of trawl-induced sediment gravity flows. The new observations reveal a highly active scenario with the nearly daily occurrence of sediment transport events linked to the passage of the

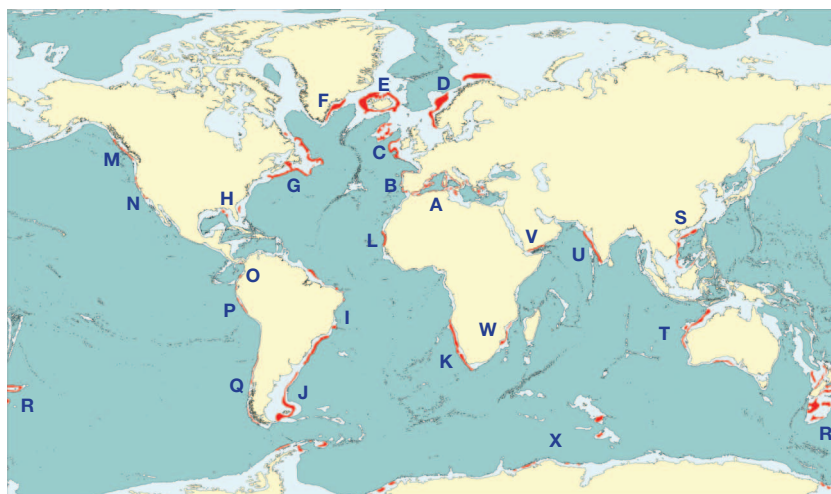


Figure 1 | Map of known trawling grounds on the world continental slopes. Red areas illustrate the extent of the trawling and the letters refer to the main fishery operating in each area, as detailed in the Supplementary Information. Darker blue regions represent water depths exceeding 2,000 m.

¹Marine Sciences Institute, CSIC, E-08003 Barcelona, Spain. ²CRG Marine Geosciences, University of Barcelona, E-08028 Barcelona, Spain.

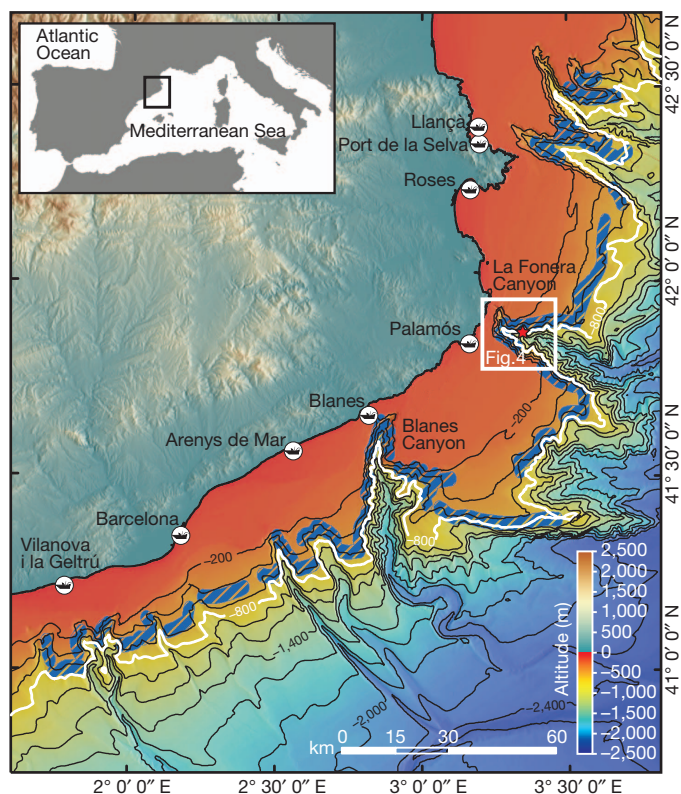


Figure 2 | Bathymetric map of the study area. Bathymetric map of the northern Catalan margin (northwestern Mediterranean) showing the main trawling grounds on the open continental slope and canyon flanks (blue hatching). Data was obtained from the VMS. The largest ports of the area (vessels in white circles) and the 800-m isobath (white thick contour) are highlighted. The location of Fig. 4 and the mooring station (red star in white box; see Fig. 3 and Supplementary Fig. 1) in La Fonera (also named Palamós) Canyon are indicated.

trawling fleet upslope of the mooring site (Fig. 3). Increases of near-bottom suspended sediment concentrations and current velocities from trawl-induced sediment gravity flows were observed repeatedly during weekdays at working hours, indicating a periodic sediment removal from fishing grounds. Such flows reached maximum downslope velocities of up to 38 cm s^{-1} and concentrations of up to 236 mg l^{-1} (Fig. 3; Supplementary Fig. 1a). Maximum instantaneous sediment fluxes at 4–6 m above the sea floor (m.a.s.f.) were $34 \text{ g m}^{-2} \text{ s}^{-1}$ and accounted for a cumulative transport of about $1.1 \text{ tonnes m}^{-2}$ during the observational period (Supplementary Fig. 1b). This near-bottom transport is of the same order of magnitude as the wintertime downslope cumulative transport induced by storms and dense shelf water cascading in the nearby Cap de Creus and Lacaze–Duthiers submarine canyon heads^{14,20}. Integrating the sediment transport from 4 to 50 m.a.s.f., a minimum of 5.4×10^3 tonnes of sediment is estimated to have been exported downslope through the monitored tributary valley in 136 days as a consequence of trawling activities (Supplementary Fig. 1c). These recurrent trawl-induced low-concentration sediment gravity flows only travel a few kilometres downslope from the canyon flank to the main canyon axis where they deposit their sediment load¹⁹. Consequently, after the industrialization of the fishing fleet, a twofold increase of the sediment accumulation rates occurred in the canyon axis¹⁸.

In 2007 a high-resolution multibeam bathymetry survey was conducted in La Fonera Canyon (Fig. 4). Analysis of the bathymetry data revealed a noticeable smoothing of bottom topography along the northern canyon flank at depths shallower than 800 m (Fig. 4a), which had been tentatively interpreted as being caused by dense shelf water cascading flowing southwards²¹. Canyon flank smoothing, however,

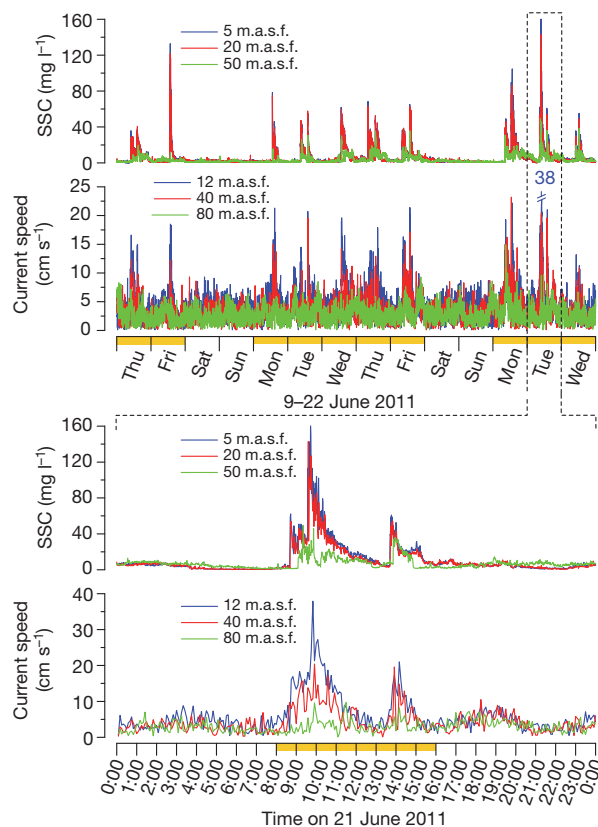


Figure 3 | Time series observations of trawling-induced sediment gravity flows. Graphs showing near-bottom suspended sediment concentration (SSC) and current speed records observed at the northern flank of La Fonera Canyon during a two-week period at different heights above the sea floor (m.a.s.f., metres above sea floor). Sediment transport events were observed during weekdays (that is, working days for the fishery), but not on weekends. The detail of a one-day record (21 June 2011) shows two sharp increases in SSC and the consequent intensification of the current speed (directed downslope towards 192°), with higher velocities closer to the sea floor that denote the occurrence of sediment gravity flows. The two peaks correspond to the two main hauls of the trawling fleet, the first one heading offshore and the second one heading to port. Orange bars on the x axes denote working days and working hours, respectively. Note the change in scale of the current-speed y axes. See mooring location in Figs 2 and 4 and Supplementary Figs 2 and 3.

was also observed locally in the southern canyon flank (Fig. 4a), away from the region potentially affected by cascading flows, challenging previous hypotheses. The fact that the lower limit of the smoothed bathymetric range coincided with the maximum trawled depths reported by the fishermen operating in the area pointed to trawling as a potential seafloor shaping agent within this depth range.

Plotting four years (2007–2010) of satellite-based navigation tracks from all large bottom trawlers operating in the area on top of the multibeam bathymetry of La Fonera Canyon revealed that navigation tracks coincided with the smoothed canyon flanks at depths shallower than 800 m (Fig. 4b) in a region characterized by homogeneous slopes (Fig. 4c) and low rugosity (a dimensionless measure of surface roughness) values (Supplementary Fig. 2). In contrast, untrawled canyon flank segments are dominated by a dense network of tributary valleys that progress upslope from the main canyon axis, reaching up to five orders of bifurcation²¹. Such complex morphology is missing in the trawled depth range and slightly below. Here the sea floor becomes smoother and only the main branches of the tributary valley networks are preserved (Fig. 4, Supplementary Fig. 3). Therefore, a causative relationship can be established between trawling-induced sediment erosion effects and the reduction of morphological complexity in the studied area. Contrasts between untrawled and trawled seafloor

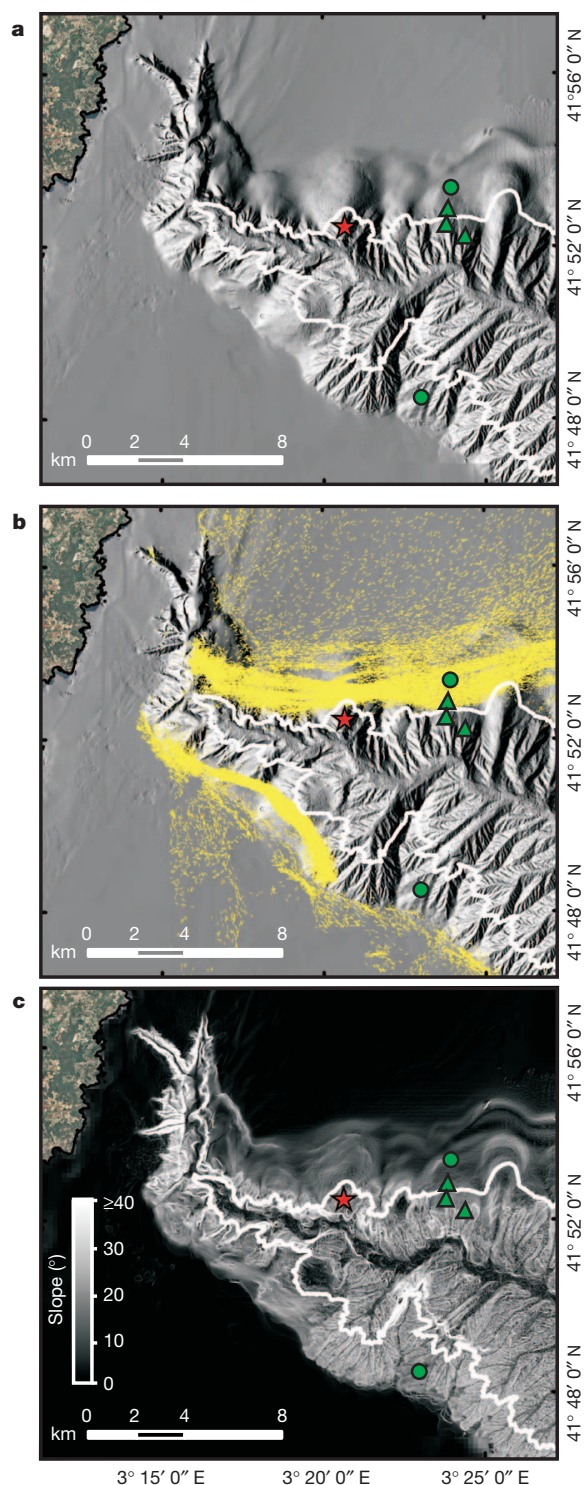


Figure 4 | Seafloor morphology and bottom trawling in La Fonera Canyon.

a, Shaded relief image from a 15-m resolution bathymetric grid showing gullied and smoothed seafloor regions, the latter at depths shallower than 800 m (white contour) especially on the northern flank but also in parts of the southern flank. **b**, Positioning data (small yellow arrows) of bottom trawling vessels fishing from 2007 to 2010. **c**, Slope map illustrating regions of homogeneous values between 10° and 30° in the most heavily trawled regions of the canyon flanks compared to untrawled regions where the gullied sea floor and the tributary valley networks are apparent, displaying higher (>40°) and lower (<10°) slopes in the walls and axes of the various valleys, respectively. The positions of the mooring station (red star; see Fig. 3 and Supplementary Fig. 1), sediment cores (green dots; see Supplementary Fig. 4) and ROV images (green triangles; see Supplementary Fig. 5) are also indicated.

regions are also evident in sediment cores (Supplementary Fig. 4) and remote operated vehicle (ROV) footage from canyon flanks (Supplementary Fig. 5). Considering only the satellite-based navigation tracks from 2010, we determine that an area of 41.6 km² was affected in the northern canyon flank alone. This represents 76.3% of the canyon flank area that is shallower than 800 m.

To provide an estimate of the volume of sediment that can be remobilized downslope by trawling activities, the integrated sediment transport during the observational period (Supplementary Fig. 1c) was converted to volume using a mean dry bulk density of 0.6 g cm⁻³ (Supplementary Fig. 4) and extrapolated to one year. Using a drainage basin of 4.2 km² (about 10% of the canyon flank trawled during 2010) for the monitored tributary valley, approximately 2.4×10^{-4} km³ is estimated to have been exported from this fishing ground annually. Assuming an equal sediment yield since the 1970s, following the industrialization of the trawling fleet¹⁸, a total of around 1.0×10^{-2} km³ has potentially been removed from this canyon flank by bottom trawling, contributing to its morphological change. The actual volume remobilized over this period could easily be larger, taking into account that the surface sediments from the fishing ground are at present over-consolidated (Supplementary Fig. 4) and that downslope sediment fluxes in the past, when the morphology was rougher and the surface sediments unconsolidated and easier to resuspend, could have been much higher. Although the estimated annual volume is similar to the annual sediment volumes involved in canyon head failures, the total estimated volume remobilized from this canyon flank over four decades is two to five orders of magnitude smaller than the sediment volumes involved in open-slope submarine failures that occur over geological periods of time (Supplementary Table 1).

Recent studies assessing the impact of anthropogenic activities in the marine ecosystems have already highlighted the effects caused by bottom trawling, both at global scale²² and in deep-sea environments, where its spatial extent has been estimated to be an order of magnitude greater than the total extent of all other anthropogenic activities²³. Given that deep-sea bottom trawling has spread around the globe during the past decades^{4,15,16} (Fig. 1), the present study raises questions over whether bottom trawling is becoming an important driver of the upper continental slope seascape evolution during the Anthropocene²⁴, at a scale that has been previously unnoticed.

The relevance of bottom trawling in the alteration of natural sediment fluxes over continental shelves has already been pointed out^{6,25}, but our results stress that, in addition to altering deep-sea sediment fluxes, bottom trawling also modifies the shape of the upper continental slope, particularly in regions that are strongly affected by this fishing technique. In these portions of the margin, the morphological complexity, as well as benthic habitat heterogeneity, has been drastically reduced, potentially affecting species diversity by regulating levels of competition, predation and physiological stress²⁶. Such seascape alteration could also affect ecosystem functioning, because deep-sea ecosystems have evolved in these environments to rely upon the original morphological complexity and biogeochemical properties of untrawled surface sediments to support benthic organisms.

This process of seascape spatial levelling has many parallels with landscape evolution under anthropogenic pressures. Changes in land use during the twentieth century resulted in a reduction of landscape diversity almost everywhere through the spreading of arable land and timber plantations, which led to the gradual disappearance of both major and minor elements of topography²⁷. Bottom trawling has been compared to forest clear-cutting²⁸, although our results suggest that a better comparison might be intensive agricultural activities. Trawled continental-slope environments are the underwater equivalent of a gullied hill slope on land, part of which has been transformed into crop fields that are ploughed regularly, thus replacing the natural contour-normal drainage pattern by levelled areas with a smaller-scale contour-parallel alignment of troughs and crests (Fig. 4, Supplementary Fig. 5). Differences exist in the processes that operate in submarine

and subaerial environments. Although farmers usually plough their land a few times per year, at sea trawling can occur on a nearly daily basis (Fig. 3, Supplementary Fig. 1). Also, agricultural activities do not generate the periodic sediment gravity flows that contribute to remove material from fishing grounds. However, the frequent repeated trawling (ploughing) over the same ground, involving displacement of sediments owing to mechanical redistribution, is what ultimately causes the levelling of the surface and produces morphological effects similar to those of a farmer's field.

Warming, sea-level rise, acidification and changes in ocean circulation have been listed as the major present and foreseeable human footprints in the ocean²⁹. This work argues for the inclusion in that list of the modification of the submarine landscape by trawling, along with the ecological and biogeochemical consequences that may derive from it. Human-induced changes on Earth's geomorphology, already well documented for emerged land³⁰, are also present in submerged regions.

METHODS SUMMARY

This study was conducted using a unique, dedicated experimental strategy combining monitoring of contemporary sediment transport processes, sediment coring, remote operated vehicle (ROV) *in situ* observations and high-resolution multibeam bathymetric surveys, coupled with the information provided by satellite-based navigation tracks of fishing vessels. Swath bathymetry data from La Fonera (also named Palamós) Canyon were obtained during the EUROLEÓN cruise onboard RV *Hespérides* in October 2007 using two Simrad multibeam echosounders: EM-1002S in shallow and EM-120 in deep waters. A bathymetric grid of 15-m resolution was generated to obtain the Digital Terrain Model used in this study. Contemporary sediment transport processes were measured by an instrumented mooring line equipped with a downward-looking 300 kHz acoustic Doppler current profiler (ADCP) from Teledyne RDI and three turbidimeters (AQUA logger 520 from the Aquatec Group) deployed from 10 May to 22 September 2011 on the northern flank of La Fonera Canyon, at 980 m depth within the Montgrí tributary valley axis. Seafloor sediment samples from the canyon flanks were collected using a six-tube multicorer from KC-Denmark during the HERMIONE-1 cruise onboard the RV *García del Cid*. A vertical slab of undisturbed sediment was removed from one of the core tubes for X-ray analysis. A second core tube was sliced at 1-cm intervals for granulometric analyses. Bottom video imagery of La Fonera canyon flanks was recorded in July 2011 during the PROMARES-OASIS DEL MAR research cruise onboard RV *Sarmiento de Gamboa* by using a high-definition Kongsberg OE14-502A camera mounted on the 2000-m-depth-rated SubAtlantic Super Mohawk 24 ROV *Liropus* belonging to the Spanish Institute of Oceanography. The satellite-based tracking of fishing trawlers operating in the study area from the Vessel Monitoring Systems (VMS) data for 2007–2010 was provided by the Fishing Monitoring Centre of the Spanish General Secretariat of Maritime Fishing (SEGEMAR).

Full Methods and any associated references are available in the online version of the paper.

Received 20 January; accepted 10 July 2012.

Published online 5 September 2012.

1. Thurstan, R. H., Brockington, S. & Roberts, C. M. The effects of 118 years of industrial fishing on UK bottom trawl fisheries. *Nature Commun.* **1**, 15 (2010).
2. Norse, E. A. *et al.* Sustainability of deep-sea fisheries. *Mar. Policy* **36**, 307–320 (2012).
3. Watling, L. & Norse, E. A. Effects of mobile fishing gear on marine benthos. *Conserv. Biol.* **12**, 1178–1179 (1998).
4. Roberts, J. M., Wheeler, A. J. & Freiwald, A. Reefs of the deep: the biology and geology of cold-water coral ecosystems. *Science* **312**, 543–547 (2006).
5. Jones, J. B. Environmental impact of trawling on the seabed: a review. *N. Z. J. Mar. Freshwat. Res.* **26**, 59–67 (1992).
6. Churchill, J. H. The effect of commercial trawling on sediment resuspension and transport over the Middle Atlantic Bight continental-shelf. *Cont. Shelf Res.* **9**, 841–865 (1989).
7. Schwinghamer, P. *et al.* Effects of experimental otter trawling on surficial sediment properties of a sandy-bottom ecosystem on the Grand Banks of Newfoundland. *Conserv. Biol.* **12**, 1215–1222 (1998).
8. Brown, E. J., Finney, B., Dommissie, M. & Hills, S. Effects of commercial otter trawling on the physical environment of the southeastern Bering Sea. *Cont. Shelf Res.* **25**, 1281–1301 (2005).
9. Nittrouer, C. A. & Wright, L. D. Transport of particles across continental shelves. *Rev. Geophys.* **32**, 85–113 (1994).

10. Pratson, L. F. *et al.* in *Continental-Margin Sedimentation: from Sediment Transport to Sequence Stratigraphy* (eds Nittrouer, C. A. *et al.*) 339–380 (Blackwell, 2007).
11. Piper, D. J. W. & Normark, W. R. Processes that initiate turbidity currents and their influence on turbidites: a marine geology perspective. *J. Sediment. Res.* **79**, 347–362 (2009).
12. Heezen, B. C., Hollister, C. D. & Ruddiman, W. F. Shaping of the continental rise by deep geostrophic contour currents. *Science* **152**, 502–508 (1966).
13. Cacchione, D. A., Pratson, L. F. & Ogston, A. S. The shaping of continental slopes by internal tides. *Science* **296**, 724–727 (2002).
14. Canals, M. *et al.* Flushing submarine canyons. *Nature* **444**, 354–357 (2006).
15. Gordon, J. D. M., Bergstad, O. A., Figueiredo, I. & Menezes, G. Deep-water fisheries of the Northeast Atlantic: I. Description and current trends. *J. Northwest Atlantic Fish. Sci.* **31**, 137–150 (2003).
16. Morato, T., Watson, R., Pitcher, T. J. & Pauly, D. Fishing down the deep. *Fish Fish.* **7**, 24–34 (2006).
17. Tobar, R. & Sardà, F. Análisis de la evolución de las capturas de gamba rosada, *Aristeus antennatus* (Risso, 1816), en los últimos decenios en Cataluña. *Inf. Téc. Inv. Pesq.* **142**, 3–20 (1987).
18. Martín, J., Puig, P., Palanques, A., Masqué, P. & García-Orellana, J. Effect of commercial trawling on the deep sedimentation in a Mediterranean submarine canyon. *Mar. Geol.* **252**, 150–155 (2008).
19. Palanques, A. *et al.* Evidence of sediment gravity flows induced by trawling in the Palamós (Fonera) submarine canyon (northwestern Mediterranean). *Deep-Sea Res.* **53**, 201–214 (2006).
20. Palanques, A. *et al.* Suspended sediment fluxes and transport processes in the Gulf of Lions submarine canyons. The role of storms and dense water cascading. *Mar. Geol.* **234**, 43–61 (2006).
21. Lastras, G. *et al.* Understanding sediment dynamics of two large submarine valleys from seafloor data: Blanes and La Fonera canyons, northwestern Mediterranean Sea. *Mar. Geol.* **280**, 20–39 (2011).
22. Halpern, B. S. *et al.* A global map of human impact on marine ecosystems. *Science* **319**, 948–952 (2008).
23. Benn, A. R. *et al.* Human activities on the deep seafloor in the North East Atlantic: an assessment of spatial extent. *PLoS ONE* **5**, e12730 (2010).
24. Crutzen, P. J. Geology of mankind. *Nature* **415**, 23 (2002).
25. Ferré, B., Durrieu de Madron, X., Estournel, C., Ulses, C. & Le Corre, G. Impact of natural (waves and currents) and anthropogenic (trawl) resuspension on the export of particulate matter to the open ocean: application to the Gulf of Lion (NW Mediterranean). *Cont. Shelf Res.* **28**, 2071–2091 (2008).
26. Levin, L. A., Sibuet, M., Gooday, A. J., Smith, C. R. & Vanreusel, A. The roles of habitat heterogeneity in generating and maintaining biodiversity on continental margins: an introduction. *Mar. Ecol.* **31**, 1–5 (2010).
27. Csorba, P. in *Anthropogenic Geomorphology* (eds Szabó, J., Dávid, L. & Dénes, L.) 39–52 (Springer, 2010).
28. Watling, L. & Norse, E. A. Disturbance of the seabed by mobile fishing gear: a comparison with forest clear-cutting. *Conserv. Biol.* **12**, 1180–1197 (1998).
29. Tyrrell, T. Anthropogenic modification of the oceans. *Phil. Trans. R. Soc. Lond. A* **369**, 887–908 (2011).
30. Price, S. J., Ford, J. R., Cooper, A. H. & Neal, C. Humans as major geological and geomorphological agents in the Anthropocene: the significance of artificial ground in Great Britain. *Phil. Trans. R. Soc. A* **369**, 1056–1084 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by various research projects: HERMIONE (226354 and CTM2010-11084-E), DOS MARES (CTM2010-21810-C03), OASIS DEL MAR—Obra Social “la Caixa”, GRACCIE-CONSOLIDER (CSD2007-00067) and REDECO (CTM2008-04973-E). We also received support from Catalan Government Grups de Recerca Consolidats grants (2009 SGR 899 and 1305). VMS data and support were provided by the Spanish General Secretariat of Maritime Fishing (SEGEMAR). Assistance at sea by the crews of RV *Hespérides*, RV *Sarmiento de Gamboa* and RV *García del Cid* is also acknowledged. J.M. was funded through a JAE-DOC contract granted by Consejo Superior de Investigaciones Científicas and co-financed by the European Social Fund. F. Sardà and T. Milligan provided comments on the manuscript. The final document benefited from a review by P. Talling.

Author Contributions All authors contributed to the design and implementation of the experimental strategy. P.P., M.C. and J.B.C. steered the integration and joint analysis of the data. P.P. and J.M. drafted the manuscript, which was critically revised and implemented by M.C. and J.B.C. M.C. designed and coordinated the acquisition of seafloor data, including multibeam bathymetry and ROV observations, and obtained VMS data. P.P., J.M. and A.P. acquired and processed the mooring time series and the sediment cores. J.M. and J.B.C. conducted the bibliographical survey of global slope fishing grounds. D.A., M.C. and P.P. produced the various maps in the paper, to which J.M. and J.B.C. contributed. G.L. coordinated the multibeam bathymetry processing and, jointly with M.C., created the slope failures table. J.B.C. provided knowledge about local fishing activities. M.C., D.A., G.L. and A.M.C. described and interpreted the whole set of seafloor data. All authors discussed the results and commented on the final version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.P. (ppuig@icm.csic.es).

METHODS

Time series observations. An instrumented mooring line equipped with a downward-looking 300 kHz acoustic Doppler current profiler (ADCP) from Teledyne RDI placed at 82 m above sea floor (m.a.s.f.) and three Seapoint turbidimeters (AQUA logger 520 from the Aquatec Group) placed at 5, 20 and 50 m.a.s.f., was deployed from 10 May to 22 September 2011 in the axis of the Montgrí tributary valley on the northern flank of La Fonera Canyon (41° 52.49' N; 3° 20.66' E), at 980 m water depth, slightly deeper than the maximum trawling depth (see position in Figs 2 and 4). Turbidimeters were programmed to measure Formazin Turbidity Units (FTU) at 1-min sampling intervals in auto-gain mode, and FTU readings were converted to estimates of suspended sediment concentration (SSC) according to ref. 31 using standard methods. The ADCP was programmed to sample at 5-min intervals with 60 pings per ensemble at minimum time ping (standard deviation 0.9 cm s^{-1}) and a depth cell size of 2 m, and recorded data until 23 September 2011. Owing to acoustic interference with the steep sea floor, the lowest cell (that is, the one closest to the bottom) with reliable current data was at 12 m.a.s.f. and the highest cell was at 80 m.a.s.f.

Sediment fluxes. To estimate the down-slope sediment transport at the mooring location, the north and east current components from each ADCP cell were transformed to along-valley and across-valley current components, using an 11° clockwise rotation from north as the orientation of the tributary valley axis. Downslope sediment fluxes were then obtained by multiplying the along-valley current velocity from each ADCP cell by the estimate of SSC at the same depth level, obtained from a linear regression between SSC measurements from 5 to 20 m.a.s.f., and from 20 to 50 m.a.s.f., averaged at 5-min intervals. Currents from 12 m.a.s.f. were assumed to be the same down to 4 m.a.s.f. to calculate near-bottom fluxes, considering SSC measurements from 5 m.a.s.f. and the currents from 4 to 6 m.a.s.f. to be the lowest cell. The time-integrated cumulative transport through each cell, computed as tonnes per square metre, was multiplied by its height (that is, 2 m) and the valley width at the water depth of the cell to obtain the mass of sediment that passed through the various levels (that is, from 4 to 50 m.a.s.f.). The integration of the transport at all levels gave an estimate of the amount of sediment that was exported downslope through the tributary valley during the observational period (136 days). This value should be considered a minimum estimate, because the flux above 50 m.a.s.f. and, more importantly, the flux from 4 m.a.s.f. to the sea floor, where most of the sediment transport probably occurs, could not be accurately calculated.

Multibeam bathymetry. Swath bathymetry data were obtained during the EUROLEÓN cruise onboard RV *Hespérides* in October 2007 using two SIMRAD multibeam echosounders: an EM-1002S for shallow depths and an EM-120 in deep waters. The EM-1002S works at a frequency of 95 kHz with 111 beams per ping, whereas the EM-120 works at a frequency of 12 kHz with 191 beams per ping. Both were operated in equidistant mode, with swath width independent of water depth and fixed to 500 m. Data were logged using Kongsberg's Seafloor Information System (SIS). Processing of both multibeam data sets was conducted using Caris HIPS and SIPS software (<http://www.caris.com/products/hips-sips/>) and included corrections for ship motion, sound speed and spurious data. A bathymetric grid of 15-m resolution was generated to obtain the final Digital Terrain Model and slope gradient map illustrated in Fig. 4, and a 50-m bathymetric grid subsample was used to perform the rugosity analysis shown in Supplementary Fig. 2.

Trawlers' satellite-based tracking. Within the framework of the Common Fisheries Policy of the European Union, the Fishing Monitoring Centre of the Spanish General Secretariat of Maritime Fishing (SEGEMAR) continuously tracks about 1,400 fishing vessels equipped with and operating satellite-based Vessel

Monitoring Systems (VMS)³², commonly referred to as 'blue boxes'. Blue boxes are standardized, protected and sealed against power cuts, the marine environment and to prevent any manipulation. Each blue box consists of a positioning unit and a satellite-based transmission-reception station, along with data logging and storage devices. VMS uses the Global Positioning System and Inmarsat-C for positioning and data transmission, respectively. Data are stored in the VMS internal memory during six months. Blue boxes also have supplementary buttons with which to select the vessel's activity at any time. The VMS provides information on a vessel's position, direction and speed and is able to determine each vessel's activity (that is, fishing, sailing, drifting or docked) at any time regardless of whether it is located in authorized or prohibited waters. The error margin in the positioning is less than 100 m and the maximum time between the transmission of a message and its reception at the Fishing Monitoring Centre is 10 min. The route followed by each vessel is known and its speed reveals whether it is sailing or fishing. For the blue and red deep-sea shrimp *Aristeus antennatus* fishery, sustained vessel speeds of less than 3.5 knots imply that the vessel is fishing. The VMS information represented in Fig. 4b corresponds to 55 different VMS identification codes (that is, vessels) for the period extending from 1 January 2007 to 31 December 2010. The VMS information in Fig. 4b was filtered by depth and speed. Vessel velocities higher than 3.5 knots and depths shallower than 130 m (that is, the mean depth of the continental shelf edge in the study area) were removed to illustrate better the actual fishing activity in slope waters. Each vessel heading has also been illustrated in Fig. 4b by means of small, course-oriented arrows.

Bottom sediment samples. Seafloor sediment samples were collected in La Fonera Canyon using a six-tube KC multicorer during the HERMIONE-1 cruise onboard the RV *García del Cid*. A vertical slab of undisturbed sediment was removed from one of the core tubes for X-radiographic analysis. A second core tube was sliced at 1 cm intervals for further analyses. Dry bulk density was determined as the ratio of freeze-dried weight to volume of wet sediment slice. For granulometric analysis, about 1 g of lyophilized sediment sample was treated with 20% H_2O_2 to remove organic matter and then dispersed by adding sodium polyphosphate solution. Grain size analysis of the prepared sample was conducted by laser diffraction using a Horiba Partica LA-950V2 particle-size analyser. The LA-950V2 detects particles in the range 0.01–3,000 μm , so sediment samples were sieved through a 2,000- μm mesh screen before analysis and the grain size distribution obtained by laser diffraction was corrected to include the relative percentage of gravel.

Remote operated vehicle imagery. Bottom video imagery of trawled and untrawled sections of La Fonera canyon flanks was recorded in July 2011 during the PROMARES-OASIS DEL MAR research cruise onboard the Spanish RV *Sarmiento de Gamboa* by using a high-definition Kongsberg OE14-502A camera mounted on the 2,000-m-depth-rated Sub-Atlantic Super Mohawk 24 ROV *Liropus* belonging to the Spanish Institute of Oceanography. Single frames were subsequently extracted from the video records using Videolan (VLC version 1.1.11 The Luggage) and Avidemux (version 2.5.4) media player software. Underwater positioning was ensured by a high-precision HiPAP 350P acoustic system linked to the DGPS positioning system of the vessel. Position accuracy of the HiPAP 350P is 0.3% of the range and its range detection accuracy is less than 20 cm.

- Guillén, J., Palanques, A., Puig, P., Durrieu de Madron, X. & Nyffeler, F. Field calibration of optical sensors for measuring suspended sediment concentration in the western Mediterranean. *Sci. Mar.* **64**, 427–435 (2000).
- Commission Regulation (EC) No. 2244/2003 of 18 December 2003 laying down detailed provisions regarding satellite-based Vessel Monitoring Systems. *Official J. Eur. Union* **L333**, 17–27 (2003); <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:333:0017:0027:EN:PDF>.

Stereocontrolled organocatalytic synthesis of prostaglandin PGF_{2α} in seven steps

Graeme Coulthard¹, William Erb¹ & Varinder K. Aggarwal¹

Prostaglandins are hormone-like chemical messengers that regulate a broad range of physiological activities, including blood circulation, digestion and reproduction^{1,2}. Their biological activities and their complex molecular architectures have made prostaglandins popular targets for synthetic organic chemists for over 40 years^{3,4}. Prostaglandin analogues are widely used as pharmaceuticals and some, such as latanoprost, which is used to treat glaucoma^{5,6}, have become billion-dollar drugs. Previously reported syntheses of these compounds are quite lengthy, and every chemical step costs time and energy, generates waste and is accompanied by material losses. Using a new bond disconnection, here we report a concise synthesis of the most complex prostaglandin, PGF_{2α}, with high levels of control of relative and absolute stereochemistry, and fewer steps. The key step is an aldol cascade reaction of succinaldehyde using proline organocatalysis to create a bicyclic enal in one step and an enantiomeric excess of 98%. This intermediate bicyclic enal is fully primed with the appropriate functionality for attachment of the remaining groups⁷. Access to this bicyclic enal will not only render existing prostaglandin-based drugs more affordable, but will also facilitate the rapid exploration of related chemical structures around the ubiquitous five-membered ring motif, such as potentially therapeutic prostaglandin analogues.

Prostaglandins, such as PGF_{2α} (Fig. 1; 1), are hormones that are responsible for the control of a myriad of essential biological processes from sleep to pain, fever to inflammation, menstruation to birth, and constriction of blood vessels to blood clotting^{1,2,8}. They are derived from arachidonic acid and are transformed by prostaglandin synthetase into a number of structurally related carbocyclic molecules. These sensitive and labile molecules are not stored in the body but are synthesized in response to stimuli. They were discovered in the early 1930s by von Euler⁹ and by the mid-1960s the structures of the first family of

prostaglandins was uncovered by Bergström *et al.* (see refs 2 and 10 for reviews). The complex structures of prostaglandins, together with their broad spectrum of biological activity, fuelled intense research activity into their synthesis, comparable to that generated from β-lactam antibiotics and steroids. Woodward¹¹, Corey¹², Stork¹³, Noyori¹⁴, Danishefsky¹⁵ and many others contributed ingenious strategies and developed new methodologies of general use in the construction of these complex molecules^{3–5}. But prostaglandins and their analogues are not only of academic interest, they have also found their way into a considerable number of pharmaceuticals. For example, an analogue of PGF_{2α} (1), latanoprost (2), is used in the treatment of glaucoma¹⁶, and achieved sales of \$1.75 billion in 2010 (ref. 17).

The manufacture of latanoprost⁶ requires 20 steps and uses the original strategy developed by Corey *et al.*¹⁸ in the synthesis of the related prostaglandin, PGF_{2α}. Corey's synthesis involved the formation of a key intermediate, the Corey lactone (3), in nine steps from cyclopentadiene (4). From this lactone, they were able to assemble the entire family of prostaglandins¹², and specifically PGF_{2α}, in eight further steps¹⁸.

In our analysis for the synthesis of prostaglandins and in common with other syntheses of PGF_{2α},^{13,15,19} we recognized that lactol 5 was an ideal late-stage intermediate because it enabled the incorporation of the upper side-chain with control of double-bond geometry through a Wittig reaction. At this point our retrosynthetic analysis departed from all previous syntheses. We were interested in disconnecting the C12–C13 bond²⁰, because we recognized that the bicyclic lactol could control the stereochemistry of the 1,4-addition of the lower side-chain onto a suitable Michael acceptor 6. By selecting an aldehyde as the electron withdrawing group of the Michael acceptor, the enal 7 could be disconnected back to a simple aldol dimerization of succinaldehyde (8), a process that could be rendered asymmetric through proline (9) catalysis^{21,22}.

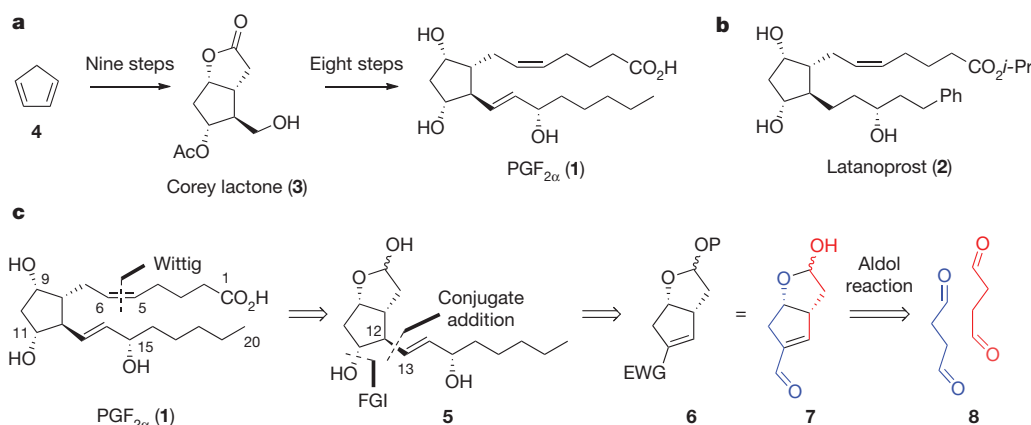


Figure 1 | Prostaglandins in nature and medicine. **a**, General schematic of Corey's synthesis of PGF_{2α}. **b**, The molecular structure of latanoprost (2), a PGF_{2α} analogue used for the treatment of glaucoma. **c**, Our retrosynthetic

analysis of PGF_{2α}. The numbered carbon atoms are referred to in the main text. FGI, functional group interconversion. Ph, phenyl. EWG, electron withdrawing group. i-Pr, isopropyl, CH(CH₃)₂.

¹School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

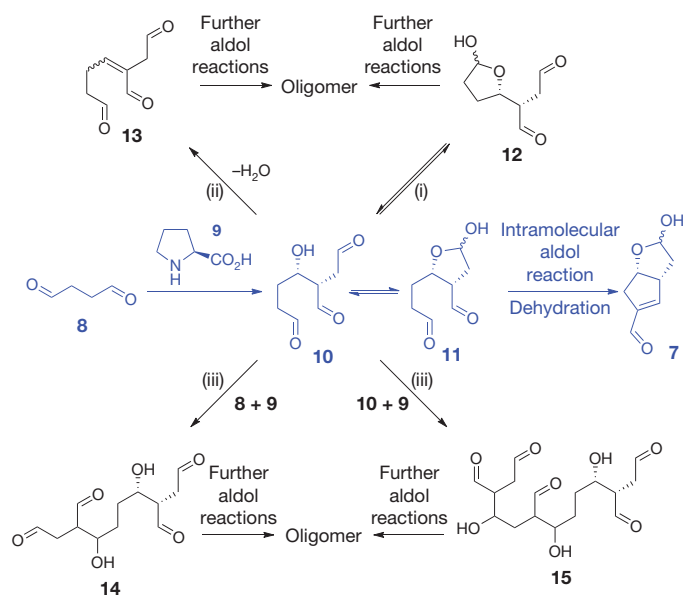


Figure 2 | Potential reaction pathways of the proline-catalysed aldol reaction of succinaldehyde. The desired pathway is shown in blue; others are labelled (i) to (iii) and described in the main text.

However, the ‘simple’ concept shown in the retrosynthesis belies a highly complex reaction cascade (blue, Fig. 2) with many potential pitfalls (black, Fig. 2): (i) the aldol product **10** is required to form the less favoured hemiacetal **11**, bearing *cis* substituents in the 5-membered ring, but not the *trans* hemiacetal **12**; (ii) the hemiacetal **11** is required to undergo an intramolecular second aldol and eliminate to give **7** but aldol **10** itself should not eliminate to give **13**; (iii) aldol **10** is a reactive trialdehyde which will be prone to undergo further aldol reactions with succinaldehyde (**8**) or with itself, leading to **14** and **15** and ultimately oligomers.

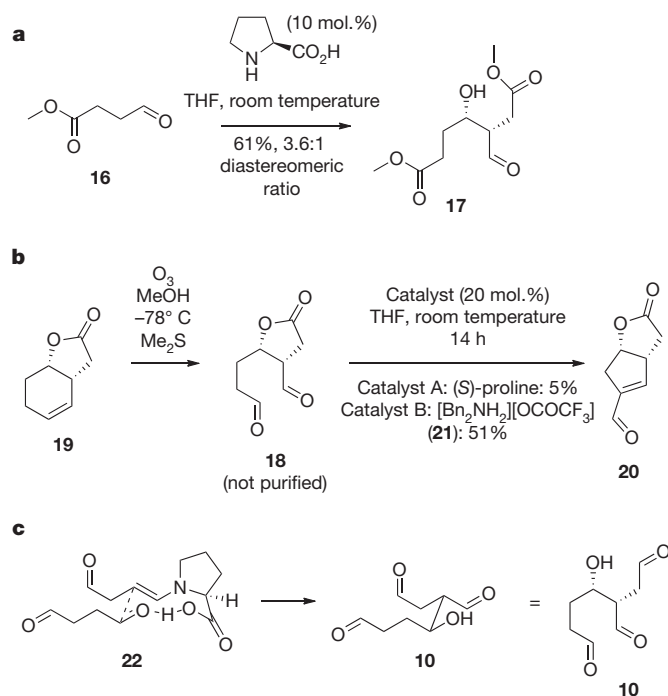


Figure 3 | Model studies. **a**, Model studies to investigate the intermolecular proline-catalysed aldol reaction of aldehydes bearing an ester at the 4-position. **b**, Model studies to investigate an intramolecular aldol reaction and dehydration. **c**, A transition state structure accounting for the observed enantioselectivity.

Perhaps unsurprisingly, the desired aldol proved highly challenging to realize in practice: treatment of succinaldehyde with proline in a range of solvents and under a variety of conditions did not deliver any of the desired bicyclic product but instead gave oligomeric material. We therefore deconstructed the reaction cascade to determine which of the two steps, the initial aldol step or the second aldol and dehydration step, was causing the problem.

Model aldehyde **16** was treated with proline (**9**) and aldol product **17** was obtained as a 3.6:1 mixture of diastereoisomers in moderate yield (Fig. 3a)²³. This showed that aldol reactions of aldehydes bearing a carbonyl group in the 4-position were suitable substrates for the proline-catalysed aldol reaction.

To test the second aldol and dehydration step, model dialdehyde **18** was prepared by ozonolysis of the known lactone **19** (Fig. 3b)²⁴. However, treatment of this dialdehyde with proline (catalyst A) provided only low conversion to the expected enal **20**, clearly indicating that the second step was the hurdle in the reaction cascade. We therefore explored alternative catalysts and found that $[\text{Bn}_2\text{NH}_2][\text{OCOCF}_3]$ (**21**, catalyst B)²⁵ was much more effective, giving enal **20** in 51% isolated yield (from **19**).

We therefore tested the aldol reaction cascade of succinaldehyde with a combination of proline and $[\text{Bn}_2\text{NH}_2][\text{OCOCF}_3]$ in a range of solvents and under a variety of conditions (see the Supplementary Information for full details). This time the reaction was successful; selected data, illustrative of the process, is summarized in Table 1. The reaction could be conducted at (unusually) low loadings of both catalysts and at relatively high concentration (2 M) (entries 1–3 in Table 1). The sequenced addition of the two catalysts and the timing was critical to the success of the reaction: if $[\text{Bn}_2\text{NH}_2][\text{OCOCF}_3]$ was present at the outset (entry 4 in Table 1), or not added at all (see earlier discussion), the reaction failed. We spent some time investigating this facet of the reaction because the low yield is a consequence of oligomerization of the intermediate trialdehyde. As the concentration of the trialdehyde builds up it is prone to undergo further aldol reactions with succinaldehyde, leading to oligomers. The second catalyst seemed to inhibit the initial proline-catalysed aldol reaction, so consecutive addition of the two catalysts was required. The optimum time to add the second catalyst was found to be dependent on the amount of proline catalyst used: using just 2% proline, the yield of the lactol **7** increased with increasing time before the addition of the second catalyst (entries 5–8 in Table 1), peaking at around 10 h. The reaction could also be conducted with just 1% proline but considerably longer reaction times were required (entry 9 in Table 1).

Table 1 | Effect of catalyst loading and time delay on yield

Entry	(S)-proline (mol.%)	Time (h)	Yield (%)
1	10	2	14
2	5	2	16
3	5	4	19
4	2	0	~2
5	2	4	10
6	2	6	16
7	2	10	20
8	2	24	20
9	1	24	18

Reactions were carried out on 200 mg of succinaldehyde **8**. ‘Time’ refers to the time before $[\text{Bn}_2\text{NH}_2][\text{OCOCF}_3]$ (2 mol.%) was added and the reaction was diluted to 1 M. The yield is the NMR (nuclear magnetic resonance) yield based on an internal standard (1,3,5-trimethoxybenzene); results are an average of five reactions.

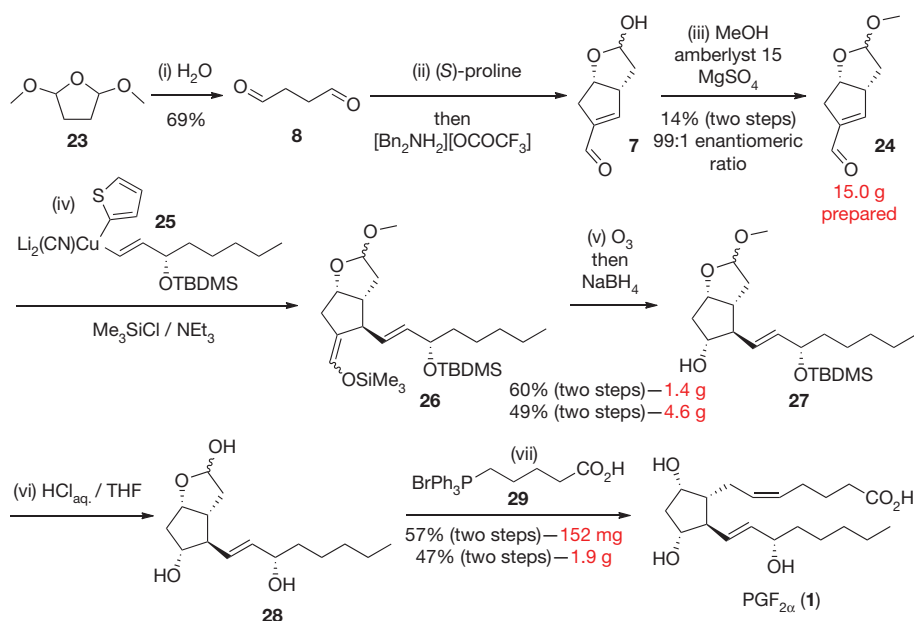


Figure 4 | A concise asymmetric synthesis of $\text{PGF}_{2\alpha}$ (**1**). Reaction conditions are as follows (see the Supplementary Information for full details). (i), H_2O , at 75°C for 4 h, then at 115°C to distil MeOH and H_2O . (ii), (S)-proline (2 mol.%), THF (2 M), at room temperature for 20 h, then $[\text{Bn}_2\text{NH}_2][\text{OCOCF}_3]$ (2 mol.%), THF (1 M), at room temperature for 14 h. (iii), MeOH (2.0 equivalents), amberlyst 15, MgSO_4 , CH_2Cl_2 , at room temperature for 14 h.

The reaction was conducted at relatively high concentration, which greatly assisted the scale-up of the reaction. Despite the low yield, work-up and purification of the enal was straightforward as the oligomeric side products could be largely removed by filtration, leaving a relatively pure crude material. Partial purification through a plug of silica gave the desired enal **7** in about 16% yield and with an enantiomeric ratio of 99:1, on the multigram scale. The absolute stereochemistry of the product was ultimately established through the synthesis of $\text{PGF}_{2\alpha}$ (**1**) and follows from the List–Houk model for this type of reaction (transition-state structure **22**, Fig. 3c)²⁶. A mixture of diastereoisomers of trialdehyde **10** was expected to be formed (a diastereoisomeric ratio of about 3.6:1 was expected from results with the model compound **16**) but as the minor diastereoisomer cannot give the alternative diastereomeric enal it must be consumed by the formation of oligomers.

The complete synthesis of $\text{PGF}_{2\alpha}$ from commercially available materials is shown in Fig. 4. Heating 2,5-dimethoxytetrahydrofuran **23** (41 g) in water followed by evaporation and extraction gave crude dialdehyde²⁷ which was directly subjected to the aldol reaction as described above. The hemi-acetal **7** was converted into a 2:1 inconsequential diastereoisomeric mixture of methoxy acetals **24** which were carried through the subsequent reaction sequence. On a large scale (285 g of **23**), further modifications were required owing to the partial decomposition of the dialdehyde during the extended time needed to remove the larger volume of solvent used for extraction. Instead, we found that after distillation of the methanol (MeOH) and most of the water, azeotropic removal of the remaining water using 2-MeTHF (2-methyltetrahydrofuran) provided a solution of the dialdehyde in a solvent (2-MeTHF) that could be used directly in the aldol reaction. Using this improved and simplified protocol, from 109.5 g of inexpensive succinaldehyde we were able to obtain 15.0 g of pure methoxy acetal **24** in 14% yield over two steps. Conjugate addition of the mixed vinyl cuprate²⁸ **25** to 4.0 g of methoxy acetal **24** followed by trapping with TMSCl (trimethylsilyl chloride) furnished the silyl enol ether **26**. Subsequent controlled ozonolysis followed by treatment with NaBH_4 gave the alcohol **27** (4.6 g, 49% over two steps). As planned, these two steps occurred with complete stereocontrol at the newly created stereogenic centres. Finally, simultaneous deprotection of the

(iv), **25** (1.1 equivalents), THF, then Me_3SiCl , Et_3N . (v), O_3 , $\text{CH}_2\text{Cl}_2/\text{MeOH}$ (3:1), -78°C , then NaBH_4 (3 equivalents), at -78°C to room temperature. (vi), 1.5% aqueous HCl/THF (3:2), at room temperature for 16 h. (vii), (4-carboxybutyl)(triphenyl)phosphonium bromide (6 equivalents), potassium *tert*-butoxide (12 equivalents), THF, at 0°C to room temperature. TBDMS, *tert*-butyldimethylsilyl.

acetal and silyl ether with aqueous HCl followed by Wittig reaction with the phosphonium salt **29** (ref. 29) gave 1.9 g of the target molecule $\text{PGF}_{2\alpha}$ (**1**), which was identical in all respects to the natural product³⁰.

Thus, we have developed a short (seven steps) synthesis of prostaglandin $\text{PGF}_{2\alpha}$ from inexpensive 2,5-dimethoxytetrahydrofuran **23**. The key step is an organocatalytic aldol dimerization reaction of succinaldehyde, which generates the bicyclic enal **7** in high enantiomeric ratio and fully primed with functionality suitable to introduce the required side chains directly. The aldol cascade required proline to perform the first aldol reaction and a second catalyst ($[\text{Bn}_2\text{NH}_2][\text{OCOCF}_3]$) to induce an intramolecular aldol reaction and elimination. Although the aldol reaction was low-yielding, the enantioselectivity was very high, isolation and purification was straightforward and the reaction could be conducted on the multigram scale. Its application in a short synthesis of the most complex of prostaglandins, $\text{PGF}_{2\alpha}$, has been demonstrated. Indeed, the bicyclic enal **7** is an ideal building block not just for the cost-effective synthesis of the whole family of prostaglandins, but for also exploring the chemical space around the ubiquitous five-membered carbocyclic ring motif, where other biologically active molecules undoubtedly lie.

METHODS SUMMARY

Procedure for formation of methoxyacetal **24.** A solution of succinaldehyde **8** (109.5 g, 1.27 mol) in 2-MeTHF (650 ml) was stirred at room temperature (20°C) (S)-proline (2.93 g, 25.4 mmol, 0.02 equivalents) was added and the reaction stirred at room temperature for 20 h. THF (650 ml) was added, followed by $[\text{Bn}_2\text{NH}_2][\text{OCOCF}_3]$ (8.43 g, 25.4 mmol, 0.02 equivalents). The reaction was stirred for a further 14 h. Celite (60 g) was added to the reaction and the volume of the reaction mixture was reduced by three quarters under reduced pressure. *tert*-Butyl methyl ether (TBME) (975 ml) was added slowly and with vigorous stirring of the mixture. The mixture was stirred for 20 min before filtration of the resulting solids. The solids were washed with TBME (3×150 ml) and the filtrate was concentrated under reduced pressure. The material was purified by column chromatography (about 600 g silica), eluting with petrol/EtOAc (6:4 to 5:5), to give the lactol **7** (as an approximately 2:1 mixture of diastereoisomers), as a brown oil.

This residue, containing **7**, was dissolved in CH_2Cl_2 (190 ml) and stirred at room temperature. MeOH (5.47 g, 6.90 ml, 170.6 mmol, 2.0 equivalents based on about 13.4% of **7** detected by internal standard in the previous reaction) was

added via syringe. Amberlyst 15 (1.22 g) and MgSO_4 (25.2 g) were added in one portion and the reaction mixture was stirred at room temperature for 14 h. The reaction mixture was filtered through a sinter funnel and the solids washed with CH_2Cl_2 (3×60 ml). The filtrate was concentrated under reduced pressure and purified by column chromatography (about 300 g silica), eluting with petrol/EtOAc (9:1 to 4:1), to give the methyl acetal **24** (as an approximately 2:1 mixture of diastereoisomers, 15.0 g, 14.0% (over two steps from succinaldehyde)) as a yellow oil. See the Supplementary Information for characterization data.

Received 9 February; accepted 11 July 2012.

Published online 15 August 2012.

- Funk, C. D. Prostaglandins and leukotrienes: advances in eicosanoid biology. *Science* **294**, 1871–1875 (2001).
- Bindra, J. S. & Bindra, R. *Prostaglandin Synthesis* (Academic Press, 1977).
- Bindra, J. S. in *The Total Synthesis of Natural Products* (ed. ApSimon, J.) Vol. 4 353–449 (Wiley, 1981).
- Das, S., Chandrasekhar, S., Yadav, J. S. & Gree, R. Recent developments in the synthesis of prostaglandins and analogues. *Chem. Rev.* **107**, 3286–3337 (2007).
- Collins, P. W. & Djuric, S. W. Synthesis of therapeutically useful prostaglandin and prostacyclin analogs. *Chem. Rev.* **93**, 1533–1564 (1993).
- Nair, S. K. & Henegar, K. E. in *Modern Drug Synthesis* (eds Li, J. J. & Johnson, D. S.) Ch. 21 329–338 (Wiley, 2010).
- Ungrin, M. D. et al. Key structural features of prostaglandin E_2 and prostanoid analogs involved in binding and activation of the human EP_1 prostanoid receptor. *Mol. Pharmacol.* **59**, 1446–1456 (2001).
- Stix, G. Better ways to target pain. *Sci. Am.* 68–71, <http://www.scientificamerican.com/article.cfm?id=better-ways-to-target-pain> (16 December 2006).
- Von Euler, U. S. Information on the pharmacological effect of natural secretions and extracts from male accessory sexual glands. *Arch. Exp. Pathol. Pharm.* **175**, 78–84 (1934).
- Flower, R. J. Prostaglandins, bioassay and inflammation. *Br. J. Pharmacol.* **147**, S182–S192 (2006).
- Woodward, R. B. et al. Novel synthesis of prostaglandin $\text{F}_{2\alpha}$. *J. Am. Chem. Soc.* **95**, 6853–6855 (1973).
- Corey, E. J. & Cheng, X.-M. *The Logic of Chemical Synthesis* (Wiley, 1995).
- Stork, G., Sher, P. M. & Chen, H. L. Radical cyclization-trapping in the synthesis of natural products. A simple, stereocontrolled route to prostaglandin $\text{F}_{2\alpha}$. *J. Am. Chem. Soc.* **108**, 6384–6385 (1986).
- Suzuki, M., Yanagisawa, A. & Noyori, R. Prostaglandin synthesis. 16. The three-component coupling synthesis of prostaglandins. *J. Am. Chem. Soc.* **110**, 4718–4726 (1988).
- Danishefsky, S. J., Paz Cabal, M. & Chow, K. Novel stereospecific silyl group transfer reactions: practical routes to the prostaglandins. *J. Am. Chem. Soc.* **111**, 3456–3457 (1989).
- Resul, B. et al. Phenyl-substituted prostaglandins: potent and selective antiglaucoma agents. *J. Med. Chem.* **36**, 243–248 (1993).
- Pfizer reports fourth-quarter and full-year 2010 results; provides 2011 financial guidance and updates 2012 financial targets. http://www.pfizer.com/files/investors/presentations/q4performance_020111.pdf (2011).
- Corey, E. J., Weinshenker, N. M., Schaaf, T. K. & Huber, W. Stereo-controlled synthesis of dl-prostaglandins $\text{F}_{2\alpha}$ and E_2 . *J. Am. Chem. Soc.* **91**, 5675–5677 (1969).
- Howard, C. C. et al. Total synthesis of prostaglandin- $\text{F}_{2\alpha}$ involving stereocontrolled and photo-induced reactions of bicyclo[3.2.0]heptanones. *J. Chem. Soc. Perkin Trans. I* 852–857 (1980).
- Corey, E. J., Nicolaou, K. C. & Beames, D. J. A short synthetic route to prostaglandins utilizing position-selective epoxide opening by the vinyl Gilman reagent. *Tetrahed. Lett.* **15**, 2439–2440 (1974).
- Mukherjee, S., Yang, J. W., Hoffmann, S. & List, B. Asymmetric enamine catalysis. *Chem. Rev.* **107**, 5471–5569 (2007).
- List, B., Lerner, R. A. & Barbas, C. F. Proline-catalyzed direct asymmetric aldol reactions. *J. Am. Chem. Soc.* **122**, 2395–2396 (2000).
- Northrup, A. B. & MacMillan, D. W. C. The first direct and enantioselective cross-aldol reaction of aldehydes. *J. Am. Chem. Soc.* **124**, 6798–6799 (2002).
- Corey, E. J. & Ravindranathan, T. A simple route to a key intermediate for the synthesis of 11-desoxyprostaglandins. *Tetrahed. Lett.* **12**, 4753–4755 (1971).
- Corey, E. J. et al. Stereospecific total synthesis of gibberellic acid. A key tricyclic intermediate. *J. Am. Chem. Soc.* **100**, 8031–8034 (1978).
- Bahmanyar, S., Houk, K. N., Martin, H. J. & List, B. Quantum mechanical predictions of the stereoselectivities of proline-catalyzed asymmetric intermolecular aldol reactions. *J. Am. Chem. Soc.* **125**, 2475–2479 (2003).
- Gourlay, B. S., Molesworth, P. P., Ryan, J. H. & Smith, J. A. A new and high yielding synthesis of unstable pyrroles via a modified Clauson-Kaas reaction. *Tetrahed. Lett.* **47**, 799–801 (2006).
- Lipshutz, B. H., Kozlowski, J. A., Parker, D. A., Nguyen, S. L. & McCarthy, K. E. More highly mixed, higher order cyanocuprates “ $\text{R}_T(2\text{-thienyl})\text{Cu}(\text{CN})\text{Li}_2$ ”. Efficient reagents which promote selective ligand transfer. *J. Organomet. Chem.* **285**, 437–447 (1985).
- de los Angeles Rey, M. et al. New synthetic strategies to vitamin D analogues modified at the side chain and D ring. Synthesis of $1\alpha,25$ -dihydroxy-16-ene-vitamin D_3 and C-20 analogues. *J. Org. Chem.* **64**, 3196–3206 (1999).
- Sheddan, N. A. & Mulzer, J. Cross metathesis as a general strategy for the synthesis of prostacyclin and prostaglandin analogues. *Org. Lett.* **8**, 3101–3104 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank EPSRC and the European Research Council (FP7/2007–2013, ERC grant no. 246785) for financial support. V.K.A. thanks the Royal Society for a Wolfson Research Merit Award and EPSRC for a Senior Research Fellowship.

Author Contributions G.C. and W.E. were involved in the discovery and subsequent development of the aldol reaction and G.C. applied it to $\text{PGF}_{2\alpha}$. V.K.A. conceived and directed the investigations and composed the manuscript with revisions provided by G.C. and W.E.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.K.A. (v.aggarwal@bristol.ac.uk).

Stabilization of Leidenfrost vapour layer by textured superhydrophobic surfaces

Ivan U. Vakarelski^{1,2}, Neelesh A. Patankar³, Jeremy O. Marston¹, Derek Y. C. Chan^{4,5} & Sigurdur T. Thoroddsen^{1,2}

In 1756, Leidenfrost¹ observed that water drops skittered on a sufficiently hot skillet, owing to levitation by an evaporative vapour film. Such films are stable only when the hot surface is above a critical temperature, and are a central phenomenon in boiling². In this so-called Leidenfrost regime, the low thermal conductivity of the vapour layer inhibits heat transfer between the hot surface and the liquid. When the temperature of the cooling surface drops below the critical temperature, the vapour film collapses and the system enters a nucleate-boiling regime, which can result in vapour explosions that are particularly detrimental in certain contexts, such as in nuclear power plants³. The presence of these vapour films can also reduce liquid–solid drag^{4–6}. Here we show how vapour film collapse can be completely suppressed at textured superhydrophobic surfaces. At a smooth hydrophobic surface, the vapour film still collapses on cooling, albeit at a reduced critical temperature, and the system switches explosively to nucleate boiling. In contrast, at textured, superhydrophobic surfaces, the vapour layer gradually relaxes until the surface is completely cooled, without exhibiting a nucleate-boiling phase. This result demonstrates that topological texture on superhydrophobic materials is critical in stabilizing the vapour layer and thus in controlling—by heat transfer—the liquid–gas phase transition at hot surfaces. This concept can potentially be applied to control other phase transitions, such as ice or frost formation^{7–9}, and to the design of low-drag surfaces at which the vapour phase is stabilized in the grooves of textures without heating¹⁰.

Heat transfer and boiling on a heated solid surface in contact with liquid are familiar phenomena observed in our daily lives, and occur in many industrial processes. As the vessel containing a liquid is progressively heated, the liquid will start to boil, with vapour bubbles forming at the hot surface. This is the nucleate-boiling regime. If the surface temperature is increased further, a continuous vapour film will form that leads to the film-boiling or Leidenfrost regime. It is known that surface properties such as roughness and wettability can alter the transition temperature from the nucleate-boiling to the film-boiling regime^{11–16}. As the low thermal conductivity of the vapour layer dramatically reduces the efficiency of heat transfer, the traditional focus is to increase the temperature at which film boiling occurs¹³. However, if the vapour layer is to be used for drag reduction^{4–6,17–20}, lowering the temperature of film boiling is beneficial, as the presence of a Leidenfrost vapour layer can reduce the hydrodynamic drag force on a heated sphere moving in liquid by up to 85%⁶.

Takata *et al.*¹² have reported that the initiation of nucleate boiling can be suppressed at what they termed “super-water-repellent” surfaces. Here we show that a textured superhydrophobic surface can eliminate the collapse of the vapour film, thus making it possible to maintain a stable vapour film at all temperatures above the boiling point of the liquid—that is, at all superheats. To this end, we investigated cooling and heating phenomena of heated surfaces that possess different degrees of hydrophobicity in water.

In our cooling experiments we use 20-mm-diameter stainless-steel spheres, and in our heating experiments 24-mm-diameter stainless-steel cylinders and stainless-steel vessels (capacity 60 ml). The stainless-steel surfaces are modified to have superhydrophilic, hydrophilic, hydrophobic or superhydrophobic wetting properties, as characterized by water-drop contact angle in air. Hydrophilic surfaces were produced by cleaning the steel surface with organic reagents, resulting in water contact angles of $<30^\circ$. Hydrophobic surfaces were produced by silanization with trichloro(1H,1H,2H,2H-perfluorooctyl)silane, giving contact angles of $\sim 100^\circ$. Superhydrophobic surfaces with water contact angles of $>160^\circ$ were produced by treatment with a commercial coating agent (Glaco Mirror Coat ‘Zero’, Soft 99 Co.) containing nanoparticles and organic reagent. Superhydrophilic surfaces, with contact angles of $<10^\circ$, were produced from the superhydrophobic surface by plasma cleaning to remove the organic coating, while retaining the same texture and geometric morphology (see Supplementary Information for details).

Figure 1a is an atomic-force micrograph of the superhydrophobic surface coating, showing its rough surface topography. The very high contact angle subtended by a superhydrophobic sphere at the air–water meniscus is shown in Fig. 1b. When the sphere is fully immersed in water, its surface has a silver, mirror-like sheen, caused by reflection from the thin air layer retained on the surface. This is a signature of the Cassie state^{21–24}, in which water is in physical contact only with the roughness peaks created by the coating (Fig. 1c). This type of air entrapment, called plastron, has been shown to reduce hydrodynamic drag on spheres moving in water²⁰, and is used by some aquatic insects to facilitate underwater breathing²⁵.

We study the cooling of superhydrophilic, hydrophilic, hydrophobic and superhydrophobic spheres that have been pre-heated in a temperature-controlled furnace and then immersed in a glass tank filled with water at 22 °C (Supplementary Fig. 4). The state of the water at the sphere surface during subsequent cooling is captured on high-speed video, and the sphere temperature is monitored by an inserted thermocouple-probe thermometer. The progress of the cooling events for the hydrophilic, hydrophobic and superhydrophobic spheres can be seen in Supplementary Movies 1, 2 and 3, respectively. The superhydrophilic and hydrophilic spheres were heated to a maximum temperature of 700 °C, but the maximum temperature to which the hydrophobic and superhydrophobic spheres could be raised was limited to $\sim 400^\circ\text{C}$ by the thermal stability of the organic coating. For the superhydrophilic sphere, the maximum temperature was not high enough to initiate film boiling; when the sphere came into contact with the water, it cooled almost instantaneously, accompanied by a vigorous release of bubbles. For all other surfaces, the initial sphere temperatures were high enough to be in the film-boiling regime, in which a continuous vapour layer coated the spheres, with a single bubble dome formed on the top of the sphere by buoyancy effects. During the initial stage of cooling, the bubble dome would grow and

¹Division of Physical Sciences and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. ²Clean Combustion Research Center, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. ³Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, B224, Evanston, Illinois 60208–3111, USA. ⁴Department of Mathematics and Statistics, University of Melbourne, Parkville 3010, Victoria, Australia. ⁵Faculty of Life and Social Sciences, Swinburne University of Technology, Hawthorn 3122, Victoria, Australia.

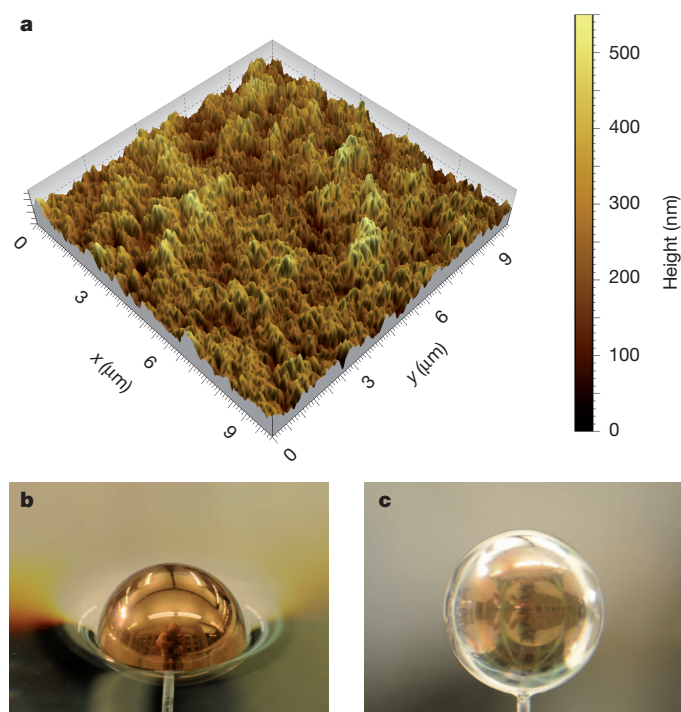


Figure 1 | Surface topography and images of the superhydrophobic sphere. **a**, Three-dimensional representation of an atomic-force micrograph of a $10\text{-}\mu\text{m}^2$ region of the superhydrophobic surface coating. The combination of high surface roughness with an organic coating provides a high water contact angle, of $>160^\circ$. **b**, Image of the superhydrophobic sphere held on the water surface, showing the high contact angle. **c**, Image of the superhydrophobic sphere immersed in water. The silver-mirror sheen of the immersed sphere is due to a thin air layer retained on the sphere surface, and is the signature appearance of a Cassie state of the water adjacent to the surface. See Supplementary Information for further details of the preparation and characterization of the superhydrophobic coating (Supplementary Figs 1, 2).

pinch off periodically (Fig. 2a). On further cooling, the film-boiling mode on the hydrophilic sphere ended with collapse of the vapour layer, marked by an explosive release of bubbles (Fig. 2c). This was followed by a short phase of rapid cooling in the nucleate-boiling mode (Fig. 2d). Film boiling on the hydrophobic sphere also ended with an almost instantaneous collapse of the vapour layer. However, for the superhydrophobic sphere no vapour-layer collapse was observed. Instead, the vapour layer gradually relaxed over the entire sphere surface, and when cooling was complete a residual bubble dome remained on the sphere apex (Fig. 2b), and the sphere surface exhibited the silver sheen indicative of a final Cassie state of the liquid adjacent to the sphere surface (Supplementary Fig. 3).

Figure 3 shows the measured sphere temperature plotted against the immersion time. The superhydrophilic sphere (Fig. 3a) begins its cooling in the nucleate-boiling regime, and its temperature drops rapidly to the pool temperature. The cooling of the hydrophilic (Fig. 3a) and hydrophobic (Fig. 3b) spheres begins in the Leidenfrost film-boiling regime. In both cases, the vapour layer collapses at the Leidenfrost point, indicated by a sharp increase in cooling rate at $\sim 420^\circ\text{C}$ for the hydrophilic sphere, and at $\sim 210^\circ\text{C}$ for the hydrophobic sphere. However, for the superhydrophobic sphere, film boiling is maintained down to the pool temperature (22°C) without vapour-layer collapse and transition to nucleate boiling (Fig. 3b).

Cooling experiments at elevated water pool temperatures of 80 and 100°C confirmed the same trends as the room-temperature cooling experiments, with respect to the sphere hydrophobicity (Supplementary Fig. 5 and Supplementary Movie 4). In fact, as long as the water contact angle at room temperature exceeded 140° , achievable at

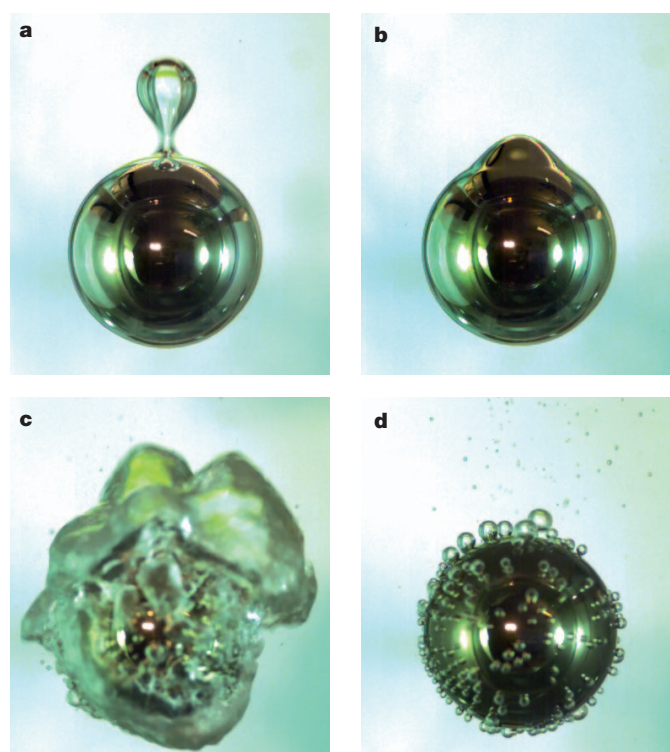


Figure 2 | High-speed camera snapshots of 20-mm steel spheres cooling in water. **a**, Bubble pinch-off from the vapour dome of a hot, superhydrophobic sphere cooling in the Leidenfrost regime. Water pool temperature is 100°C ; sphere temperature is $T_s = 200^\circ\text{C}$. **b**, Final state of the same superhydrophobic sphere, cooled to the pool temperature ($T_s = 100^\circ\text{C}$). **c**, **d**, Hydrophilic sphere cooling in water at 100°C , at the moment of the explosive transition from film boiling to nucleate boiling (**c**, $T_s = 275^\circ\text{C}$) and during nucleate boiling (**d**, $T_s = 200^\circ\text{C}$). Sphere cooling events in water at 22°C are shown in Supplementary Movies 1 (hydrophilic sphere), 2 (hydrophobic sphere) and 3 (superhydrophobic sphere). Supplementary Movie 4 shows hydrophilic and superhydrophobic spheres cooling in water at 100°C .

textured superhydrophobic surfaces, no Leidenfrost transition was observed (Supplementary Figs 7, 8).

We performed heating experiments to further characterize surface properties at the boiling transition. A cartridge heater placed inside a stainless-steel cylinder (Fig. 4a and Supplementary Fig. 9) was immersed in a water-filled beaker sitting on a hot plate that heated the water to slow boiling, $T_w = 100^\circ\text{C}$. The surface temperature of the cylinder, T_s , was measured with an inserted thermocouple at constant heat flux, by controlling the applied power to the heater. We acquired high-speed video sequences along with the temperature measurements.

Figure 4d shows the measured variations of heat flux with surface temperature at steady state for the four types of surface wettability. As expected for the low values of heat flux used here, the heat exchange occurred in the nucleate-boiling regime (Fig. 4b and Supplementary Fig. 13) for the superhydrophilic and hydrophilic surfaces. However, as in the sphere-cooling experiments, the superhydrophobic surface was able to sustain a vapour layer in the Leidenfrost regime for all surface superheats. This is readily observed in both the appearance of the vapour layer and the magnitude of the increase in surface temperature with increasing heat flux (Fig. 4d).

When the temperature of the superhydrophobic cylinder was in equilibrium with the ambient water ($T_s = T_w = 100^\circ\text{C}$), the cylinder's surface appeared smooth and shiny. With increasing heater power, the appearance of surface ripples⁶ accompanied a gradual transition to the Leidenfrost regime (Fig. 4c, Supplementary Fig. 10 and Supplementary Movie 5). Additional experiments with the immersion heater (see Supplementary Fig. 11) showed that for the superhydrophobic surface there was no hysteresis between heating and cooling experiments.

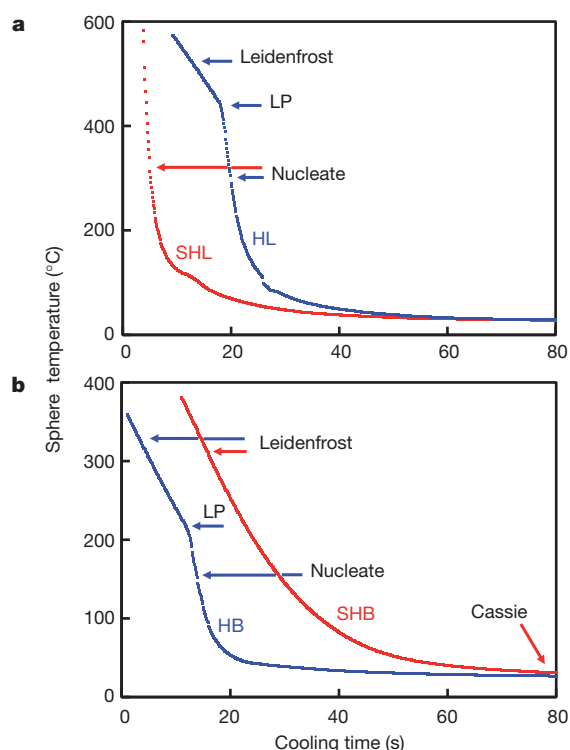


Figure 3 | Sphere temperature versus cooling time for 20-mm steel spheres held in water at 22 °C. **a**, Superhydrophilic sphere (contact angle $<10^\circ$, SHL) and hydrophilic sphere (contact angle $\sim 30^\circ$, HL). **b**, Hydrophobic sphere (contact angle $\sim 100^\circ$, HB) and superhydrophobic surface sphere (contact angle $>160^\circ$, SHB). Arrows indicate the Leidenfrost film-boiling regime, Leidenfrost-point transition (LP) and nucleate-boiling regime. The Leidenfrost transition in the superhydrophilic case occurs above the maximum temperature shown in the figure. In the superhydrophobic case, the Leidenfrost state relaxes to the Cassie state without passing through the nucleate-boiling regime.

Keeping the surface immersed in water for up to 24 hours did not alter the smooth transition to the Leidenfrost regime (Supplementary Fig. 11). Although in such cases the surfaces appeared matte instead of shiny, suggesting that the vapour layer was not in a pure Cassie state, on heating, they returned smoothly to the Leidenfrost state (Supplementary Fig. 12). The same recovery property of superhydrophobic surfaces has been demonstrated recently using electrophoresis²⁶.

For the hydrophobic surface, it was possible to maintain the Leidenfrost vapour regime only if the cylinder was overheated before immersion to $\sim 350^\circ\text{C}$, and the heater power was maximized on immersion. When the surface temperature was allowed to fall below $\sim 170^\circ\text{C}$, the Leidenfrost vapour layer collapsed, and the switched to the nucleate-boiling branch (Fig. 4d).

Experiments with a stainless-steel cup placed on a hot plate showed the same behaviour as with the immersion heater (Supplementary Figs 14, 15). The good agreement among the cooling experiments and the two types of heating experiment confirm the geometry-independence and universality of the superhydrophobic surface effect in stabilizing the Leidenfrost vapour layer (compare Fig. 4d with Supplementary Figs 6 and 15).

By combining results for cooling experiments at constant initial heat content and high initial temperatures, and heating experiments at constant heating power but over a range of lower superheat temperatures, we have shown that the mode of heat transfer at the solid-liquid interface can be controlled by a combination of the surface morphology and surface chemistry of the solid surface. The observations for the hydrophilic and the hydrophobic surfaces are consistent with the standard transition between nucleate and film boiling²⁷. The strong

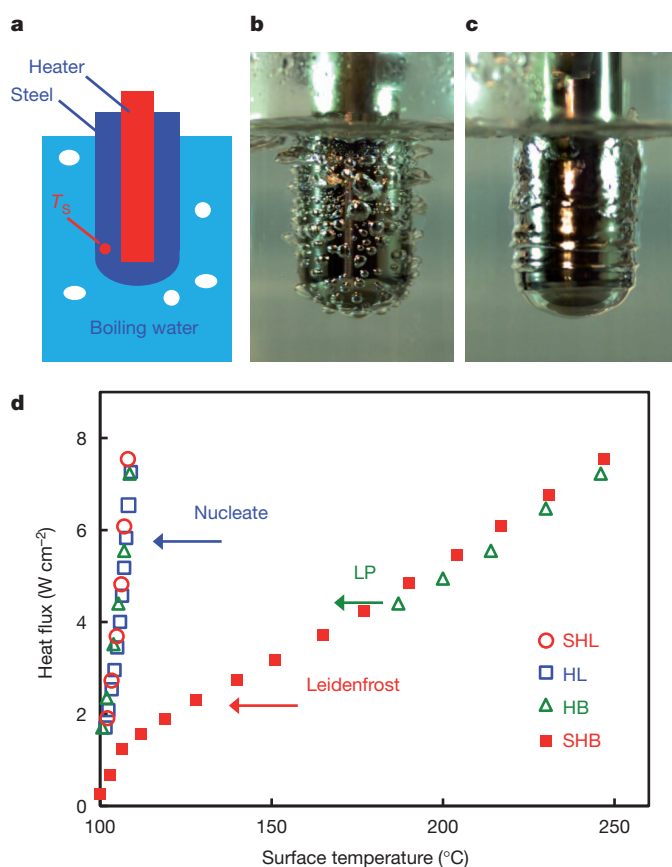


Figure 4 | Surface temperature versus heat flux in heating experiments.

a, Schematics of the immersion heater. **b**, **c**, High-speed camera snapshots of the heater in boiling (100°C) water for the case of **(b)** hydrophobic surface in the nucleate-boiling regime, at surface temperature $T_s = 106^\circ\text{C}$; and **(c)** superhydrophobic surface in the Leidenfrost regime, $T_s = 210^\circ\text{C}$. **d**, Dependence of the heat flux on the heater surface temperature for four types of surface in the nucleate-boiling and Leidenfrost film-boiling regimes. Abbreviations as in Fig. 3.

effect of the superhydrophilic surface in increasing the Leidenfrost temperature and enhancing boiling heat transfer has also been investigated recently¹³. However, we now show that for textured superhydrophobic surfaces the heat exchange always occurs in the presence of a Leidenfrost vapour layer, within the film-boiling regime. This behaviour has not been observed previously and represents a significant departure from the characteristic form of the classical boiling heat-transfer curve^{10,28,29} of heat flux versus surface superheat temperature. The currently accepted form of the boiling curve comprises a critical heat-flux maximum and a Leidenfrost-point minimum that characterizes the transition between nucleate and film boiling. But at a textured superhydrophobic surface, the heat flux is a smoothly increasing function of superheat temperature, owing to the vapour layer that is always maintained at such surfaces (Supplementary Fig. 17).

The vapour layer on the superhydrophobic surface is sustained by the high surface roughness and porosity, coupled with the intrinsic, non-wetting hydrophobicity due to surface chemical treatment. These are the requisite conditions for the existence of a large non-wetting vapour layer (see Supplementary Information discussion on stabilization of vapour layers due to texture, and Supplementary Figs 18, 19). Furthermore, the three-dimensional nature of the surface roughness permits only small areas of direct contact between the peaks of the rough surface and the liquid interface. Contributions to the total heat transfer by such small areas are thus limited, and any onset of heterogeneous nucleation of bubbles will simply be taken up in the surface vapour layer that in effect will contribute to a smooth transition to the Leidenfrost regime. These conclusions are in complete accord with

what we found when repeating Leidenfrost's experiments¹ of drop evaporation at heated surfaces. For hydrophobic surfaces we observed the classical transition from nucleate to film-boiling mode, whereas at superhydrophobic surfaces nucleate boiling is absent at all superheats (Supplementary Fig. 16).

Although it has been demonstrated that the stability of the vapour layer in Leidenfrost drop experiments is sensitive to external perturbations such as vibrations, cooling rate and internal fluid circulation within the drops³¹, and the hypothesis has been advanced that when the fluid drop contact angle at the surface attains the limiting value of 180°, the film-boiling regime might be possible at all superheat temperatures³², this study has demonstrated that with the combination of surface roughness and chemical modifications Leidenfrost vapour layer stability and the film-boiling behaviour are readily achievable and remain robust under general operational conditions.

With the recent development of a wide variety of laboratory-designed and commercially available textured surface coatings with superhydrophobic^{23,24}, superamphiphobic^{8,9}, anti-frost⁷ or switchable-hydrophobicity³⁰ properties, the effect on thermal exchange shown here opens possibilities for new applications of such coatings, ranging from the design of efficient heat-exchange devices to technologies for aqueous drag reduction.

METHODS SUMMARY

In cooling experiments we used 20-mm-diameter stainless-steel spheres (FRITSCH GmbH). In heating experiments we used an immersion heater or a 60-ml stainless-steel cup. The immersion heater is a 24-mm-diameter stainless-steel hollow cylinder with an electric heater cartridge fitted inside. All stainless-steel surfaces are modified as follows. Smooth hydrophilic surfaces are produced by cleaning with organic reagents. Smooth hydrophobic surfaces are produced by silanization with trichloro(1H,1H,2H,2H-perfluorooctyl)silane. Textured superhydrophobic surfaces are produced by treatment with a commercial superhydrophobic coating agent (Glaco Mirror Coat 'Zero', Soft 99 Co.). Textured superhydrophilic surfaces are produced by plasma cleaning of the surfaces previously treated with the superhydrophobic agent. A 2-mm hole is drilled into the spheres, the immersion-heater cylinder wall or the bottom of the cup vessel to allow the insertion of a thermocouple probe to measure the surface temperature. In cooling experiments, the heated sphere is quenched in a water-filled vessel and the sphere temperature is recorded as function of time. In heating experiments, the cylindrical heater is immersed in a water-filled glass beaker placed on a hot plate to control the water temperature, and the heater surface temperature is measured as a function of the applied power to the heater cartridge in the cylinder. The heat transfer process on the sphere surface or the immersion-heater surface is recorded using a high-speed video camera (Photron Fastcam SA-5). In the cup vessel experiments, the water-filled cup is placed on a hot plate and the vessel surface temperature is measured as a function of the hot-plate temperature. The same setup is used to measure the evaporation time of water droplets placed on the bottom of the cup. Further details of methods are available in Supplementary Information.

Received 11 February; accepted 18 July 2012.

- Leidenfrost, J. G. *De aquae communis nonnullis qualitatibus tractatus* (Duisburg, 1756); transl. Wares, C. On the fixation of water in diverse fire. *Int. J. Heat Mass Transfer* **9**, 1153–1166 (1966).
- Bernardin, J. D. & Mudawar, I. The Leidenfrost point: experimental study and assessment of existing models. *Trans. Am. Soc. Mech. Eng.* **121**, 894–903 (1999).
- Berthoud, G. Vapor explosions. *Annu. Rev. Fluid Mech.* **32**, 573–611 (2000).
- Linke, H. *et al.* Self-propelled Leidenfrost droplets. *Phys. Rev. Lett.* **96**, 154502 (2006).
- Lagubeau, N., Le Merrer, M., Clanet, C. & Quéré, D. Leidenfrost on a ratchet. *Nature Phys.* **7**, 395–398 (2011).
- Vakarelski, I. U., Marston, J. O., Chan, D. Y. C. & Thoroddsen, S. T. Drag reduction by Leidenfrost vapor layers. *Phys. Rev. Lett.* **106**, 214501 (2011).
- Mishchenko, L. *et al.* Design of ice-free nanostructured surfaces based on repulsion of impacting water droplets. *ACS Nano* **4**, 7699–7707 (2010).
- Wong, T. S. *et al.* Bioinspired self-repairing slippery surfaces with pressure-stable omniphobicity. *Nature* **477**, 443–447 (2011).
- Deng, X., Mammen, L., Butt, H. J. & Vollmer, D. Candle soot as a template for a transparent robust superamphiphobic coating. *Science* **335**, 67–70 (2012).
- Patankar, N. A. Supernucleating surfaces for nucleate boiling and dropwise condensation heat transfer. *Soft Matter* **6**, 1613–1620 (2010).
- Wang, C. H. & Dhir, V. K. Effect of surface wettability on active nucleation site density during pool boiling of water on a vertical surface. *J. Heat Transfer* **115**, 659–669 (1993).
- Takata, Y., Hidaka, S. & Kohno, M. in *Proc. Fifth Intl Conf. Enhanced, Compact and Ultra-compact Heat Exchangers: Science, Engineering and Technology* (eds Shah, R. K. *et al.*) 300–304 (Engineering Conferences International, 2005).
- Chen, R. *et al.* Nanowires for enhanced boiling heat transfer. *Nano Lett.* **9**, 548–553 (2009).
- Liu, G. & Craig, V. S. J. Macroscopically flat and smooth superhydrophobic surfaces: heating induced wetting transitions up to the Leidenfrost temperature. *Faraday Discuss.* **146**, 141–151 (2010).
- Liu, G., Fu, L., Rode, A. V. & Craig, V. S. J. Water droplet motion control on superhydrophobic surfaces: exploiting the Wenzel-to-Cassie transition. *Langmuir* **27**, 2595–2600 (2011).
- Kim, H., Truong, B., Buongiorno, J. & Hu, L. On the effect of surface roughness height, wettability, and nanoporosity on Leidenfrost phenomena. *Appl. Phys. Lett.* **98**, 083121 (2011).
- Zvirin, Y., Hewitt, G. R. & Kenning, D. B. R. Boiling on free-falling spheres: drag and heat transfer coefficients. *Exp. Heat Transf.* **3**, 185–214 (1990).
- Rothstein, J. P. Slip on superhydrophobic surfaces. *Annu. Rev. Fluid Mech.* **42**, 89–109 (2010).
- Ceccio, S. L. Friction drag reduction of external flows with bubble and gas injection. *Annu. Rev. Fluid Mech.* **42**, 183–203 (2010).
- McHale, G., Newton, M. I. & Shirtcliffe, N. J. Immersed superhydrophobic surfaces: gas exchange, slip and drag reduction properties. *Soft Matter* **6**, 714–719 (2010).
- Cassie, A. B. D. & Baxter, S. Wettability of porous surfaces. *Trans. Faraday Soc.* **40**, 0546–0550 (1944).
- Patankar, N. A. On the modeling of hydrophobic contact angles on rough surfaces. *Langmuir* **19**, 1249–1253 (2003).
- Larmour, I. A., Bell, S. E. J. & Saunders, G. S. Remarkably simple fabrication of superhydrophobic surfaces using electroless galvanic deposition. *Angew. Chem. Int. Edn Engl.* **46**, 1710–1712 (2007).
- Quéré, D. Wetting and roughness. *Annu. Rev. Mater. Res.* **38**, 71–99 (2008).
- Flynn, M. R. & Bush, J. W. M. Underwater breathing: the mechanics of plastron respiration. *J. Fluid Mech.* **608**, 275–296 (2008).
- Lee, C. & Kim, C.-J. Underwater restoration and retention of gases on superhydrophobic surfaces for drag reduction. *Phys. Rev. Lett.* **106**, 014502 (2011).
- Dhir, V. K. & Purohit, G. P. Subcooled film boiling heat transfer from spheres. *Nucl. Eng. Des.* **47**, 49–66 (1978).
- Nukiyama, S. Maximum and minimum values of heat Q transmitted from metal to boiling water under atmospheric pressure. *J. Jpn Soc. Mech. Engrs* **37**, 367–374 (1934).
- Witte, L. C. & Lienhard, J. H. On the existence of two “transition” boiling curves. *Int. J. Heat Mass Transf.* **25**, 771–779 (1982).
- Celestini, F. & Kirstetter, G. Effect of an electric field on a Leidenfrost droplet. *Soft Matter* **8**, 5992–5995 (2012).
- Baumeister, K. J., Hendricks, R. C. & Hamill, T. D. *Metastable Leidenfrost States* (NASA Technical Note D3226, 1966).
- Carey, V. P. in *Liquid-Vapor Phase-Change Phenomena* 2nd edn 353–356 (Taylor and Francis, 2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the KAUST Machine Workshop, KAUST Electronics Workshop and G. D. Li for assistance in setting up the cooling and heating experiments, and L. Chen and B. Chew from KAUST Advanced Nanofabrication, Imaging and Characterization Core Lab facilities for assistance in AFM and SEM imaging characterization of the superhydrophobic coating.

Author Contributions I.U.V. conceived research and designed the experiments. I.U.V. and J.O.M. carried out the experiments. N.A.P., D.Y.C.C. and S.T.T. contributed with discussions, analysis and theoretical interpretation of the experimental results. I.U.V., D.Y.C.C. and N.A.P. wrote the manuscript. All authors edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.U.V. (ivanuriev.vakarelski@kaust.edu.sa) or N.A.P. (n-patankar@northwestern.edu).

A 61-million-person experiment in social influence and political mobilization

Robert M. Bond¹, Christopher J. Fariss¹, Jason J. Jones², Adam D. I. Kramer³, Cameron Marlow³, Jaime E. Settle¹ & James H. Fowler^{1,4}

Human behaviour is thought to spread through face-to-face social networks, but it is difficult to identify social influence effects in observational studies^{9–13}, and it is unknown whether online social networks operate in the same way^{14–19}. Here we report results from a randomized controlled trial of political mobilization messages delivered to 61 million Facebook users during the 2010 US congressional elections. The results show that the messages directly influenced political self-expression, information seeking and real-world voting behaviour of millions of people. Furthermore, the messages not only influenced the users who received them but also the users' friends, and friends of friends. The effect of social transmission on real-world voting was greater than the direct effect of the messages themselves, and nearly all the transmission occurred between 'close friends' who were more likely to have a face-to-face relationship. These results suggest that strong ties are instrumental for spreading both online and real-world behaviour in human social networks.

Recent experimental studies^{6,14–16} have attempted to measure the causal effect of social influence online. At the same time, there is increasing interest in the ability to use online social networks to study and influence real-world behaviour^{17–19}. However, online social networks are also made up of many 'weak-tie' relationships²⁰ that may not facilitate social influence²¹, and some studies suggest that online communication may not be an effective medium for influence²². An open question is whether online networks, which harness social information from face-to-face networks, can be used effectively to increase the likelihood of behaviour change and social contagion.

One behaviour that has been proposed to spread through networks is the act of voting in national elections. Voter turnout is significantly correlated among friends, family members and co-workers in observational studies^{23,24}. Voter mobilization efforts are effective at increasing turnout²⁵, particularly those conducted face-to-face and those that appeal to social pressure²⁶ and social identity²⁷. There is also evidence from one face-to-face field experiment that voting is 'contagious', in the sense that mobilization can spread from person to person within two-person households²⁸. Although anecdotal accounts suggest that online mobilization has made a big difference in recent elections²¹, a meta-analysis of email experiments suggests that online appeals to vote are ineffective²⁴.

Voter mobilization experiments^{26–28} have shown that most methods of contacting potential voters have small effects (if any) on turnout rates, ranging from 1% to 10%. However, the ability to reach large populations online means that even small effects could yield behaviour changes for millions of people. Furthermore, as many elections are competitive, these changes could affect electoral outcomes. For example, in the 2000 US presidential election, George Bush beat Al Gore in Florida by 537 votes (less than 0.01% of votes cast in Florida). Had Gore won Florida, he would have won the election.

To test the hypothesis that political behaviour can spread through an online social network, we conducted a randomized controlled trial

with all users of at least 18 years of age in the United States who accessed the Facebook website on 2 November 2010, the day of the US congressional elections. Users were randomly assigned to a 'social message' group, an 'informational message' group or a control group. The social message group ($n = 60,055,176$) was shown a statement at the top of their 'News Feed'. This message encouraged the user to vote, provided a link to find local polling places, showed a clickable button reading 'I Voted', showed a counter indicating how many other Facebook users had previously reported voting, and displayed up to six small randomly selected 'profile pictures' of the user's Facebook friends who had already clicked the I Voted button (Fig. 1). The informational message group ($n = 611,044$) was shown the message, poll information, counter and button, but they were not shown any faces of friends. The control group ($n = 613,096$) did not receive any message at the top of their News Feed.

The design of the experiment allowed us to assess the impact that the treatments had on three user actions; clicking the I Voted button, clicking the polling-place link and voting in the election. Clicking the I Voted button is similar to traditional measures of self-reported voting, but here users reported their vote to their social community rather than to a researcher. We therefore use this action to measure political self-expression, as it is likely to be affected by the extent to which a user desires to be seen as a voter by others. In contrast, social desirability should not affect other user actions in the same way. Clicking the polling-place link took users to a separate website that helped them to find a polling location, and this action was not reported to the user's social community. We therefore use this action to measure a user's desire to seek information about the election. Finally, we used a group-level process to study the validated voting behaviour of 6.3 million users matched to publicly available voter records (see Supplementary Information).

We first analyse direct effects. We cannot compare the treatment groups with the control group to assess the effect of the treatment on self-expression and information seeking, because the control group did not have the option to click an I Voted button or click on a polling-place link. However, we can compare the proportion of users between the two treatment groups to estimate the causal effect of seeing the faces of friends who have identified themselves as voters (Fig. 1). Users who received the social message were 2.08% (s.e.m., 0.05%; t -test, $P < 0.01$) more likely to click on the I Voted button than those who received the informational message (20.04% in the social message group versus 17.96% in the informational message group). Users who received the social message were also 0.26% (s.e.m., 0.02%; $P < 0.01$) more likely to click the polling-place information link than users who received the informational message (Fig. 1).

Although acts of political self-expression and information seeking are important in their own right, they do not necessarily guarantee that a particular user will actually vote. As such, we also measured the effect that the experimental treatment had on validated voting, through examination of public voting records. The results show that users

¹Political Science Department, University of California, San Diego, La Jolla, California 92093, USA. ²Psychology Department, University of California, San Diego, La Jolla, California 92093, USA. ³Data Science, Facebook, Inc., Menlo Park, California 94025, USA. ⁴Medical Genetics Division, University of California, San Diego, La Jolla, California 92093, USA.

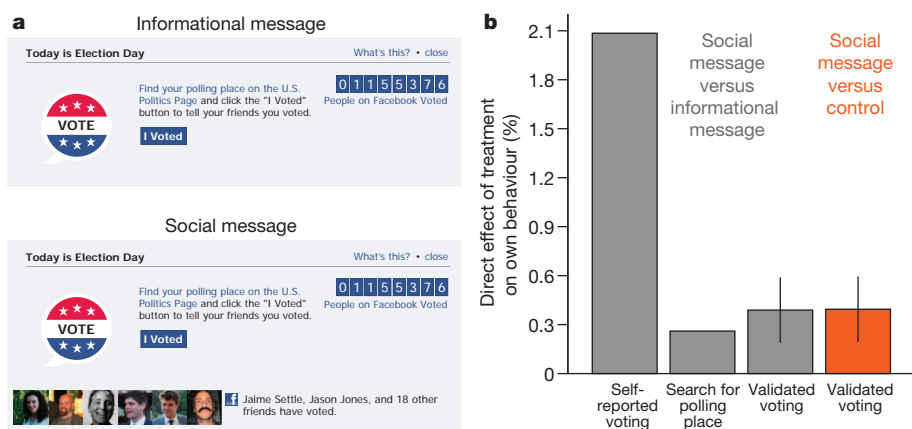


Figure 1 | The experiment and direct effects. **a, b**, Examples of the informational message and social message Facebook treatments (**a**) and their direct effect on voting behaviour (**b**). Vertical lines indicate s.e.m. (they are too small to be seen for the first two bars).

who received the social message were 0.39% (s.e.m., 0.17%; *t*-test, $P = 0.02$) more likely to vote than users who received no message at all. Similarly, the difference in voting between those who received the social message and those who received the informational message was 0.39% (s.e.m., 0.17%; *t*-test, $P = 0.02$), suggesting that seeing faces of friends significantly contributed to the overall effect of the message on real-world voting. In fact, turnout among those who received the informational message was identical to turnout among those in the control group (treatment effect 0.00%, s.e.m., 0.28%; $P = 0.98$), which raises doubts about the effectiveness of information-only appeals to vote in this context.

These results show that online political mobilization can have a direct effect on political self-expression, information seeking and real-world voting behaviour, and that messages including cues from an individual's social network are more effective than information-only appeals. But what about indirect effects that spread from person to person in the social network? Users in our sample had on average 149 Facebook friends, with whom they share social information, although many of these relationships constitute 'weak ties'. Past research indicates that close friends have a stronger behavioural effect on each other than do acquaintances or strangers^{9,11,13,21}. We therefore expected mobilization to spread more effectively online through 'strong ties'.

To distinguish users who are likely to have close relationships, we used the degree to which Facebook friends interacted with each other on the site (see Supplementary Information for more detail). Higher levels of interaction indicate that friends are more likely to be physically proximate and suggest a higher level of commitment to the friendship, more positive affect between the friends, and a desire for the friendship to be socially recognized²⁹. We counted the number of interactions between each pair of friends and categorized them by decile, ranking them from the lowest to highest percentage of interactions. A validation study (see Supplementary Information) shows that friends in the highest decile are those most likely to be close friends in real life (Fig. 2a).

We then used these categories to estimate the effect of the mobilization message on a user's friends. Random assignment means that any relationship between the message a user receives and a friend's behaviour is not due to shared attributes, as these attributes are not correlated with the treatment (see Supplementary Information). To measure a per-friend treatment effect, we compared behaviour in the friends connected to a user who received the social message to behaviour in the friends connected to a user in the control group. To account for dependencies in the network, we simulate the null distribution using a network permutation method (see the Supplementary Information). Monte Carlo simulations suggest that this method minimizes the risk of false positives and recovers true causal effects without bias (see Supplementary Information).

Figure 2 shows that the observed per-friend treatment effects increase as tie-strength increases. All of the observed treatment effects fall outside the null distribution for expressed vote (Fig. 2b), suggesting that they are significantly different from chance outcomes. For validated vote (Fig. 2c), the observed treatment effect is near zero for weak ties, but it spikes upwards and falls outside the null distribution for the top two deciles. This suggests that strong ties are important for the spread of real-world voting behaviour. Finally, the treatment effect for polling place search gradually increases (Fig. 2d), with several of the effects falling outside the 95% confidence interval of the null distribution.

To simplify the analysis and reporting of results, we arbitrarily define 'close friends' as people who were in the eightieth percentile or higher (decile 9) of frequency of interaction among all friendships in the sample (see the Supplementary Information). 'Friends' are all other Facebook friends who had less interaction. A total of 60,491,898 (98%) users in our sample had at least 1 close friend, with the average user having about 10 close friends (compared with an average of 139 friends who were not close).

The results suggest that users were about 0.011% (95% confidence interval (CI) of null distribution -0.009% to 0.010%) more likely to engage in an act of political self-expression by clicking on the I Voted button than they would have been had their friend seen no message. Similarly, for each close friend who received the social message, an individual was on average 0.099% (null 95% CI -0.042% to 0.048%) more likely to express voting.

We also found an effect in the validated vote sample. For each close friend who received the social message, a user was 0.224% (null 95% CI -0.181% to 0.174%) more likely to vote than they would have been had their close friend received no message. Similarly, for information-seeking behaviour we found that for each close friend who received the social message, a user was 0.012% (null 95% CI -0.012% to 0.012%) more likely to click the link to find their polling place than they would have been had their close friends received no message. In both cases there was no evidence that other friends had an effect (see Supplementary Information). Thus, ordinary Facebook friends may affect online expressive behaviour, but they do not seem to affect private or real-world political behaviours. In contrast, close friends seem to have influenced all three.

The magnitude of these contagion effects are small per friend, but it is important to remember that they result from a single message, and in many cases it was not possible to change the target's behaviour. For example, users may have already voted by absentee ballot before Election Day, or they may have logged in to Facebook too late to vote or to influence other users' voting behaviour. In other words, all effects measured here are intent-to-treat effects rather than treatment-on-treated effects, which would be greater if we had better information about who was eligible to receive the treatment.

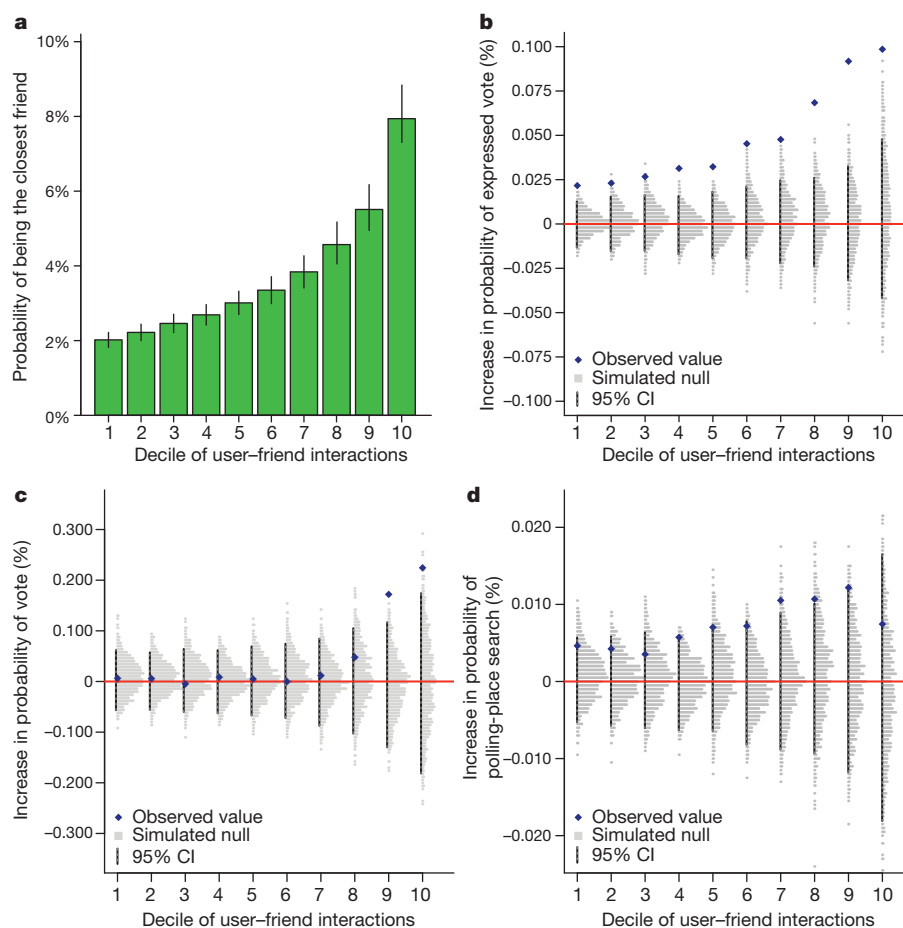


Figure 2 | The effect of mobilization treatment that a friend received on a user's behaviour. **a–d,** A validation study shows that at increasing levels of interaction, Facebook friends are more likely to have a close real-world relationship (**a**; see also the Supplementary Information). As the interaction increases, so does the observed per-friend effect of friend's treatment on a user's

expressed voting (**b**), validated voting (**c**) and polling-place search (**d**). Blue diamonds indicate the observed treatment effect. Horizontal grey bars show the null distribution derived from simulations of identical networks in which the topology and incidence of the behaviour and treatment are the same but the assignments of treatment are randomly reassigned.

Moreover, the scale of the number of users, their friendship connections and the potential voters in a given election is very large. We estimated the per-user effect (the per-friend effect multiplied by the average number of friends per user) and the total effect (the per-user effect multiplied by the total number of users) on the behaviour of everyone in the sample (see Supplementary Information). The results suggest that friends generated an additional 886,000 expressed votes (+1.4%, null 95% CI –1.1% to 1.1%), and close friends generated a further 559,000 votes (+0.9%, null 95% CI –0.3% to 0.3%). In the Supplementary Information we also show that close friends of close friends (2 degrees of separation) generated an additional 1 million expressed votes (+1.7%, null 95% CI –0.8% to 0.9%). Thus, the treatment clearly had a significant impact on political self-expression and how it spread through the network, and even weak ties seem to be relevant to its spread.

However, the effect of the social message on real-world validated vote behaviour and polling-place search was more focused. The results suggest that close friends generated an additional 282,000 validated votes (+1.8%, null 95% CI –1.3% to 1.2%) and an additional 74,000 polling-place searches (+0.1%, null 95% CI –0.1% to 0.1%), but there is no evidence that ordinary friends had any effect on either of these two behaviours. In other words, close friendships accounted for all of the significant contagion of these behaviours, in spite of the fact that they make up only 7% of all friendships on Facebook.

To put these results in context, it is important to note that turnout has been steadily increasing in recent US midterm elections, from 36.3% of the voting age population in 2002 to 37.2% in 2006, and to

37.8% in 2010. Our results suggest that the Facebook social message increased turnout directly by about 60,000 voters and indirectly through social contagion by another 280,000 voters, for a total of 340,000 additional votes. That represents about 0.14% of the voting age population of about 236 million in 2010. However, this estimate does not include the effect of the treatment on Facebook users who were registered to vote but who we could not match because of nicknames, typographical errors, and so on. It would be complex to estimate the number of users on Facebook who are in the voter record but unmatchable, and it is not clear whether treatment effects would be of the same magnitude for these individuals, so we restrict our estimate to the matched group that we were able to sample and observe. This means it is possible that more of the 0.60% growth in turnout between 2006 and 2010 might have been caused by a single message on Facebook.

The results of this study have many implications. First and foremost, online political mobilization works. It induces political self-expression, but it also induces information gathering and real, validated voter turnout. Although previous research suggested that online messages do not work¹⁹, it is possible that conventional sample sizes may not be large enough to detect the modest effect sizes shown here. We also show that social mobilization in online networks is significantly more effective than informational mobilization alone. Showing familiar faces to users can dramatically improve the effectiveness of a mobilization message.

Beyond the direct effects of online mobilization, we show the importance of social influence for effecting behaviour change. Our

validation study shows that close friends exerted about four times more influence on the total number of validated voters mobilized than the message itself. These results are similar to those from a prior network simulation study based on observational data that suggested each act of voting on average generates an additional three votes as this behaviour spreads through the network³⁰. Thus, efforts to influence behaviour should pay close attention not only to the effect a message will have on those who receive it but also to the likelihood that the message and the behaviour it spurs will spread from person to person through the social network. And, in contrast to the results for close friends, we find that Facebook friends have less effect. Online mobilization works because it primarily spreads through strong-tie networks that probably exist offline but have an online representation. In fact, it is plausible that unobserved face-to-face interactions account for at least some of the social influence that we observed in this experiment.

More broadly, the results suggest that online messages might influence a variety of offline behaviours, and this has implications for our understanding of the role of online social media in society. Experiments are expensive and have limited external validity, but the growing availability of cheap and large-scale online social network data¹⁷ means that these experiments can be easily conducted in the field. If we want to truly understand—and improve—our society, wellbeing and the world around us, it will be important to use these methods to identify which real world behaviours are amenable to online interventions.

Received 17 April; accepted 18 July 2012.

- Berger, J. & Le Mens, G. How adoption speed affects the abandonment of cultural tastes. *Proc. Natl Acad. Sci. USA* **106**, 8146–8150 (2009).
- Borgatti, S. P. *et al.* Network analysis in the social sciences. *Science* **323**, 892–895 (2009).
- Ohtsuki, H., Hauert, C., Lieberman, E. & Nowak, M. A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502–505 (2006).
- Onnela, J. P. & Reed-Tsochias, F. Spontaneous emergence of social influence in online systems. *Proc. Natl Acad. Sci. USA* **107**, 18375–18380 (2010).
- Apicella, C. L., Marlowe, F. W., Fowler, J. H. & Christakis, N. A. Social networks and cooperation in hunter-gatherers. *Nature* **481**, 497–501 (2012).
- Aral, S. & Walker, D. Identifying influential and susceptible members of social networks. *Science* **337**, 337–341 (2012).
- Centola, D. An experimental study of homophily in the adoption of health behavior. *Science* **334**, 1269–1272 (2011).
- Gonzalez-Bailon, S., Borge-Holthoefer, J., Rivero, A. & Moreno, Y. The dynamics of protest recruitment through an online network. *Sci. Rep.* **1**, 197 (2011).
- Christakis, N. A. & Fowler, J. H. Social contagion theory: examining dynamic social networks and human behavior. Preprint at <http://arXiv.org/abs/1109.5235v2> (2011).
- Aral, S., Muchnik, L. & Sundararajan, A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl Acad. Sci. USA* **106**, 21544–21549 (2009).
- Christakis, N. A. & Fowler, J. H. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370–379 (2007).
- Fowler, J. H. & Christakis, N. A. The dynamic spread of happiness in a large social network. *Br. Med. J.* **337**, a2338 (2008).
- Christakis, N. A. & Fowler, J. H. The collective dynamics of smoking in a large social network. *N. Engl. J. Med.* **358**, 2249–2258 (2008).
- Salganik, M. J., Dodds, P. S. & Watts, D. J. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* **311**, 854–856 (2006).
- Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
- Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *Proc. Natl Acad. Sci. USA* **107**, 5334–5338 (2010).
- Lazer, D. *et al.* Computational social science. *Science* **323**, 721–723 (2009).
- Ellison, N. B., Steinfield, C. & Lampe, C. The benefits of facebook “friends”: Social capital and college students’ use of online social network sites. *J. Comput. Mediat. Commun.* **12**, 1143–1168 (2007).
- Traud, A. L., Kelsic, E. D., Mucha, P. J. & Porter, M. A. Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.* **53**, 526–543 (2011).
- Granovetter, M. The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
- Christakis, N. A. & Fowler, J. H. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives* (Little, Brown, and Company, 2009).
- Nickerson, D. W. Does email boost turnout? *Q. J. Polit. Sci.* **2**, 369–379 (2007).
- Kenny, C. B. Political participation and effects from the social environment. *Am. J. Pol. Sci.* **36**, 259–267 (1992).
- Huckfeldt, R. & Sprague, J. *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign* (Cambridge Univ. Press, 1995).
- Gerber, A. S. & Green, D. P. Does canvassing increase voter turnout? A field experiment. *Proc. Natl Acad. Sci. USA* **96**, 10939–10942 (1999).
- Gerber, A. S., Green, D. P. & Larimer, C. W. Social pressure and voter turnout: evidence from a large-scale field experiment. *Am. Polit. Sci. Rev.* **102**, 33–48 (2008).
- Bryan, C. J., Walton, G. M., Rogers, T. & Dweck, C. S. Motivating voter turnout by invoking the self. *Proc. Natl Acad. Sci. USA* **108**, 12653–12656 (2011).
- Nickerson, D. W. Is voting contagious? Evidence from two field experiments. *Am. Polit. Sci. Rev.* **102**, 49–57 (2008).
- Vitak, J. *et al.* It’s complicated: Facebook users’ political participation in the 2008 election. *Cyberpsychol. Behav. Soc. Netw.* **14**, 107–114 (2011).
- Fowler, J. H. in *The Social Logic of Politics: Personal Networks as Contexts for Political Behavior* (ed. Zuckerman, A.) 269–287 (Temple Univ. Press, 2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Aral, J. Berger, M. Cebrian, D. Centola, N. Christakis, C. Dawes, L. Gee, D. Green, C. Kam, P. Loewen, P. Mucha, J. P. Onnela, M. Porter, O. Smirnov and C. Volden for comments on early drafts. This work was supported in part by the James S. McDonnell Foundation, and the University of Notre Dame and the John Templeton Foundation as part of the Science of Generosity Initiative.

Author Contributions All authors contributed to study design, data collection, analysis and preparation of the manuscript. J.H.F. secured funding.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.F. (jhfowler@ucsd.edu).

ERRATUM

doi:10.1038/nature11423

Erratum: Non-invasive prenatal measurement of the fetal genome

H. Christina Fan, Wei Gu, Jianbin Wang, Yair J. Blumenfeld, Yasser Y. El-Sayed & Stephen R. Quake

Nature **487**, 320–324 (2012); doi:10.1038/nature11251

In this Article, the statement of the competing financial interests was inadvertently omitted in the print version. The statement was corrected online on 18 July 2012.

ERRATUM

doi:10.1038/nature11425

Erratum: An origin of the radio jet in M87 at the location of the central black hole

Kazuhiro Hada, Akihiro Doi, Motoki Kino, Hiroshi Nagai,
Yoshiaki Hagiwara & Noriyuki Kawaguchi

Nature **477**, 185–187 (2011); doi:10.1038/nature10387

In Fig. 3 of this Letter, the axes labels were inadvertently exchanged. The y -axis label should read “Declination offset from 43-GHz core (mas)” and the x -axis label should read: “RA offset from 43-GHz core (mas)”. The PDF and HTML versions have been corrected online.

Impact of caloric restriction on health and survival in rhesus monkeys from the NIA study

Julie A. Mattison¹, George S. Roth², T. Mark Beasley³, Edward M. Tilmont¹, April M. Handy^{1,4}, Richard L. Herbert⁵, Dan L. Longo⁶, David B. Allison⁷, Jennifer E. Young¹, Mark Bryant⁸, Dennis Barnard⁹, Walter F. Ward¹⁰, Wenbo Qi¹¹, Donald K. Ingram¹² & Rafael de Cabo¹³

Calorie restriction (CR), a reduction of 10–40% in intake of a nutritious diet, is often reported as the most robust non-genetic mechanism to extend lifespan and healthspan. CR is frequently used as a tool to understand mechanisms behind ageing and age-associated diseases. In addition to and independently of increasing lifespan, CR has been reported to delay or prevent the occurrence of many chronic diseases in a variety of animals. Beneficial effects of CR on outcomes such as immune function^{1,2}, motor coordination³ and resistance to sarcopenia⁴ in rhesus monkeys have recently been reported. We report here that a CR regimen implemented in young and older age rhesus monkeys at the National Institute on Aging (NIA) has not improved survival outcomes. Our findings contrast with an ongoing study at the Wisconsin National Primate Research Center (WNPRC), which reported improved survival associated with 30% CR initiated in adult rhesus monkeys (7–14 years)⁵ and a preliminary report with a small number of CR monkeys⁶. Over the years, both NIA and WNPRC have extensively documented beneficial health effects of CR in these two apparently parallel studies. The implications of the WNPRC findings were important as they extended CR findings beyond the laboratory rodent and to a long-lived primate. Our study suggests a separation between health effects, morbidity and mortality, and similar to what has been shown in rodents^{7–9}, study design, husbandry and diet composition may strongly affect the life-prolonging effect of CR in a long-lived nonhuman primate.

For over 20 years, the NIA has studied the effects of CR in long-lived nonhuman primates (NHPs) (*Macaca mulatta*, average lifespan in captivity is ~27 years and maximum reported lifespan is ~40 years), to verify whether the life-prolonging effects observed in lower organisms also occur in monkeys and thus, might plausibly translate to human ageing^{10,11}. The NIA CR study began in 1987 at the NIH Animal Center¹². CR was initiated in monkeys of varying ages to evaluate the impact of age of onset of CR on its biological effects. Study design has been reported elsewhere^{12,13}. Male and female monkeys were enrolled into the study at young, middle and older ages¹². Data reported here are grouped as either young-onset (includes juvenile, adolescent and adult) or old-onset monkeys. Supplementary Table 1 reports the current census.

Any animal that died underwent a complete necropsy by a board-certified pathologist. A gross description of the pathology related to each organ was provided along with the probable cause of death and any contributing factors. Survival data were analysed in two ways:

all-cause mortality and age-related deaths; a distinction also reported previously⁵. In both studies (NIA and WNPRC), age-related survival excluded deaths due to acute conditions that do not have an age-related increase in risk such as gastrointestinal bloat, anaesthesia, injury or endometriosis. Pathology details are in the Supplementary Information.

Old-onset CR monkeys (16–23 years) did not live longer than controls in either the all-cause (Fig. 1a) or age-related survival analysis (there were three cases of non-age related deaths in the CR group and 2 in the control group, graph not shown). In this group, males had significantly longer survival compared to females ($P = 0.0003$) and neither sex benefitted from CR. To date, four CR monkeys and one control from the old-onset group have lived beyond 40 years. Although CR has not increased mean or maximum lifespan relative to control, 50% survival for the females is 27.8 years and 35.4 years for the males, exceeding the ~27 year median lifespan previously reported for monkeys in captivity¹⁴. These monkeys may have benefitted from excellent husbandry conditions and thus CR started at older ages provided no additional increase in survival. Furthermore, there were no apparent differences in causes of death between the two diet groups. Neoplasia, cardiovascular disease, amyloidosis and general organism deterioration in the oldest animals were equally represented in both diet groups.

Old-onset CR was beneficial on several measures of metabolic health and overall function. Both male and female CR monkeys weighed less than the control counterparts, although the diet effect was greater in the males. In longitudinal measures from serum of fasted monkeys, triglycerides, cholesterol and glucose levels increased with age for both male and female controls. However, triglycerides were significantly lower in the CR monkeys ($F_{(1,21)} = 5.76$, $P = 0.026$) (Fig. 1b), and cholesterol remained significantly lower in the CR males (Fig. 1c) ($F_{(40,774)} = 1.53$, $P = 0.02$). At the oldest ages, fasting glucose was numerically lower in the CR monkeys (Fig. 1d) and significantly lower in CR males compared to controls ($P = 0.04$). On a single measure of plasma-free isoprostane, an indicator of oxidative stress, control males had significantly higher levels than the CR monkeys (23.24 ± 1.25 versus 15.93 ± 1.97 pg ml⁻¹; $P = 0.009$). In contrast, we previously reported that old-onset CR may negatively affect immune function¹⁵.

Current survival curves for the young-onset male and females are shown in Fig. 2a (all-cause mortality) and Fig. 2b (age-related mortality). No significant diet effects are noted in survival between control and CR

¹Laboratory of Experimental Gerontology, National Institute on Aging, NIH Animal Center, 16701 Elmer School Road Building 103, Dickerson, Maryland 20842, USA. ²GeroScience, 1124 Ridge Road Pylesville, Maryland 21132, USA. ³Department of Biostatistics, Ryals Public Health Bldg 343C University of Alabama at Birmingham, 1530 3rd Avenue S, Birmingham, Alabama 35294, USA. ⁴SoBran, Inc., 4000 Blackburn Lane, Suite 100, Burtonsville, Maryland 20866, USA. ⁵National Institute of Allergy and Infectious Disease, NIH Animal Center, 16701 Elmer School Road, Building 102, Dickerson, Maryland 20842, USA. ⁶Laboratory of Molecular Biology and Immunology, National Institute on Aging, NIH, 251 Bayview Boulevard Room 08C228, Baltimore, Maryland 21224, USA. ⁷Office of Energetics, University of Alabama at Birmingham, 1665 University Boulevard, RPH 140J Birmingham, Alabama 35294, USA. ⁸Office of the Director, Diagnostic and Research Services Branch, NIH, Bldg 28A, Room 114, 28 Service Road West, Bethesda, Maryland 20814, USA. ⁹Office of the Director, Diagnostic and Research Services Branch, NIH, Building 14A, Room 119A, 14 Service Road West, Bethesda, Maryland 20814, USA. ¹⁰Department of Physiology/Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78229, USA. ¹¹Department of Cellular and Structural Biology/Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78229, USA. ¹²Nutritional Neuroscience and Aging Laboratory, Pennington Biomedical Research Center, Louisiana State University, 6400 Perkins Road, Baton Rouge, Louisiana 70808, USA. ¹³Laboratory of Experimental Gerontology, National Institute on Aging, NIH, 251 Bayview Boulevard Suite 100, Baltimore, Maryland 21224, USA.

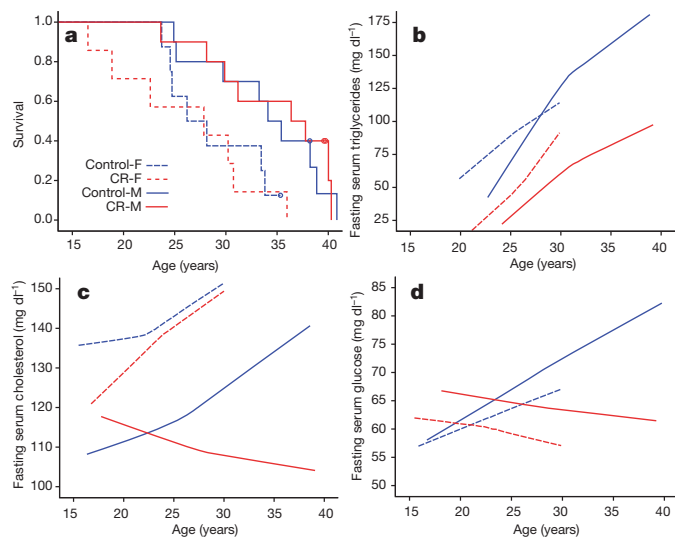


Figure 1 | Survival curve and triglycerides, cholesterol and glucose levels for old-onset monkeys. **a**, Kaplan–Meier survival curve for all-cause mortality for old-onset monkeys. All-cause mortality was analysed using Cox regression with age of onset, sex and diet as predictors. The effect of diet was not significant ($P = 0.934$) and sex was the only significant predictor ($P = 0.003$). Open circles represent alive monkeys. **b**, Fasting serum triglycerides (mg dl^{-1}) predicted from age-dependent individual-specific trajectories for old-onset monkeys. Triglyceride levels increased with age ($F_{(16,162)} = 2.12$, $P = 0.0096$) and CR monkeys had significantly lower levels than control ($F_{(1,21)} = 5.76$, $P = 0.026$). F, female; M, male. Overall triglyceride trajectories were based on 243 observations for 34 monkeys (50 observations for 8 control-F; 81 for 10 control-M; 32 for 7 CR-F; 80 for 9 CR-M). Age breakdowns for all figures are in the supplementary material. **c**, Cholesterol predicted from age-dependent individual-specific trajectories for old-onset monkeys. Cholesterol levels increased with age ($F_{(53,774)} = 1.54$, $P = 0.009$), and male monkeys had significantly lower levels than females ($F_{(1,24)} = 23.60$, $P < 0.0001$). A significant three-way diet–sex–age interaction ($F_{(40,774)} = 1.53$, $P = 0.02$) indicated that cholesterol levels increased with age for control males whereas CR males tended to have a slight reduction in cholesterol. Thus, at older ages (>30 years), CR male monkeys have significantly lower cholesterol levels compared to controls. Overall cholesterol trajectories were based on 994 observations for 28 animals (204 for 7 control-F; 301 for 7 control-M; 134 for 5 CR-F; and 355 for 9 CR-Male). **d**, Fasting serum glucose (mg dl^{-1}) levels predicted from age-dependent individual-specific trajectories for old-onset monkeys. Five glucose measurements above 100 mg dl^{-1} for one diabetic control-M were omitted to remove the influence of these outliers on the analyses and graphs. There were significant changes in glucose over time ($F_{(20,285)} = 10.48$, $P < 0.0001$) and males and females were significantly different in the trends over time ($F_{(18,285)} = 3.58$, $P < 0.0001$) with males having increases in glucose levels over time, whereas the glucose levels of the females slightly decreased. The overall CR difference was not significant, $F_{(1,22)} = 1.18$, $P = 0.288$, and the CR differences in trend over time were not significant, $F_{(20,285)} = 1.23$, $P = 0.2259$. Additional analyses stratified by sex conditions showed that control males had significantly higher glucose levels compared to CR males, $F_{(1,14)} = 5.27$, $P = 0.04$. Overall glucose trajectories were based on 387 observations for 34 monkeys (79 observations for 8 control-F; 131 for 10 control-M; 48 for 7 CR-F; 129 for 9 CR-M).

monkeys for either analysis. Statistical controls are described in Methods. Of the original 86 monkeys in the young-onset cohorts, 24% (11/46) of the control animals and 20% (8/40) of the CR group died of age-related causes. The NIA findings contrast with the adult-onset study at WNPRC that demonstrated a beneficial CR effect in which 37% of the control monkeys had died from age-related causes compared to only 13% in the CR group. When accounting for all deaths, the trend persisted with 9 control and 13 CR animals dying of non-age related causes. Survival probabilities for all NIA age groups combined are shown in Supplementary Fig. 1a, b.

Considering that just less than 50% of young monkeys are still alive, these data do not represent final lifespan curves in this study. On

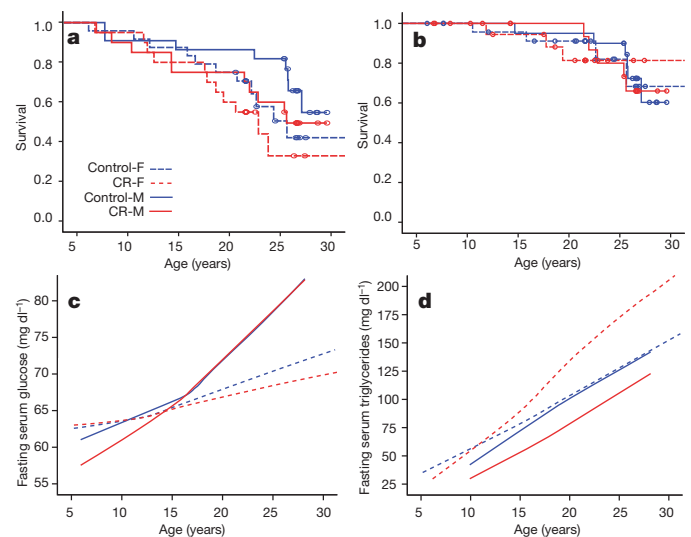


Figure 2 | Survival curves and glucose and triglycerides levels for young-onset monkeys. **a**, **b**, Kaplan–Meier survival curves for all-cause (**a**) and age-related mortality (**b**) for young-onset monkeys. Both were analysed using Cox regression with age of onset, origin, sex and diet ($P = 0.255$ and $P = 0.975$, respectively) as predictors with none of these factors being statistically significant. Open circles represent monkeys that are still alive and non-age related deaths in **b**. **c**, Fasting serum glucose (mg dl^{-1}) levels predicted from age-dependent individual-specific trajectories for young-onset monkeys. 14 glucose measurements above 100 mg dl^{-1} in diabetic monkeys (7 observations for 3 control-M; 5 for 2 CR-M; 2 for 1 control-F) were omitted to remove the influence of these outliers on the analyses and graphs. There were significant changes in glucose over time ($F_{(18,1112)} = 11.24$, $P < 0.0001$), and males and females were significantly different in the trends over time ($F_{(18,1112)} = 1.98$, $P = 0.0088$) with males having a larger increase in glucose levels over time. There was no significant difference due to diet group. Overall glucose trajectories were based on 1,260 observations for 81 monkeys (346 observations for 23 control-F; 350 for 20 control-M; 281 for 20 CR-F; 283 for 18 CR-M). **d**, Fasting serum triglycerides (mg dl^{-1}) predicted from age-dependent individual-specific trajectories for young-onset monkeys. There were significant changes in triglycerides over time ($F_{(14,843)} = 17.59$, $P < 0.0001$) and males and females were significantly different in the trends over time ($F_{(14,843)} = 5.36$, $P < 0.0001$). Furthermore, there was a diet–sex interaction indicating that the overall effect of CR on triglycerides was significantly different for male and female monkeys ($F_{(1,68)} = 5.07$, $P = 0.0276$). Specifically, CR males had lower triglycerides than control males. By contrast, CR females had higher triglyceride levels than control females. Overall triglyceride trajectories were based on 973 observations for 81 monkeys (266 observations for 23 control-F; 280 for 20 control-M; 213 for 20 CR-F; 214 for 18 CR-M).

the basis of lifespan projections using the hazard function¹⁶, most animals are projected to be dead 10 years from now and the estimated probability statistics indicates a likelihood of less than 0.1% chance that the overall survival outcome would favour the CR group. The probability that a significantly different effect on mean survival will emerge in the next 5–10 years of the study is very low; however, a potential effect on maximum lifespan cannot be ruled out.

As there is a clear difference in CR effect on mortality between the colonies at NIA and WNPRC, further comparisons of these two longitudinal studies are warranted and planned. In an estimate of NIA's current data (as of 1 December 2011) to the published WNPRC data summarized as of 22 February 2008 and reported in ref. 5, NIA monkeys, both control and CR, may have a lifespan advantage comparable to the WNPRC CR monkeys.

Although they eat less (Supplementary Table 2) and weigh less¹³, young-onset CR monkeys lack many of the expected CR benefits. Fasting serum glucose levels were not significantly lower in the CR monkeys compared to control (Fig. 2c), and only the CR males had somewhat lower triglycerides compared to respective controls ($P = 0.051$) (Fig. 2d). However, in a ligature-induced model of

inflammation in the oral cavity¹, we have shown an improved immune response in young-onset CR monkeys and beneficial effects in T cells isolated from adolescent-onset males².

The incidence of cancer was markedly improved in young-onset CR monkeys ($P = 0.028$ compared to controls); in fact, neoplasia has not been identified in any monkey from this group (Fig. 3a). In contrast, five of the six cases in young-onset control monkeys were considered the cause of death with a mean age at diagnosis of 22.8 ± 1.7 years. Glucoregulatory function was also improved in CR monkeys (Fig. 3a). However, two cases of diabetes have been diagnosed in CR monkeys; thus, the prevention of obesity did not prevent the occurrence of insulin-dependent diabetes and further investigation of the aetiology of such cases is of interest. Interestingly, CR did not reduce the incidence of cardiovascular disease as was reported in the WNPRC colony. Our findings are based on tissue pathology because these diagnoses were identified after death.

An analysis of first occurrence of age-related disease was done on the NIA monkeys using the same disease criteria as defined by the WNPRC study. These conditions included: cancer, diabetes, arthritis, diverticulosis and cardiovascular disease. Although age-related diseases

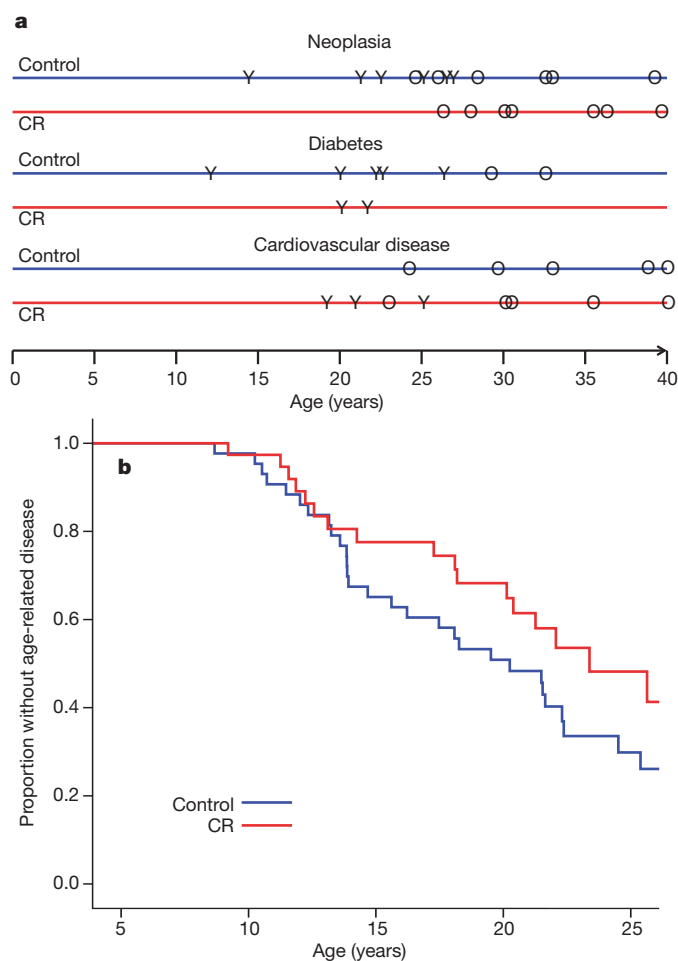


Figure 3 | Incidence and estimated proportions of age-related diseases. **a**, Incidence of three major age-related conditions. Age at diagnosis is represented with a 'Y' for young-onset monkeys and 'O' for old-onset monkeys. Animals may be represented more than once if multiple conditions existed. **b**, Estimated proportions for the first occurrence of any age-related disease in each monkey from the young-onset age group (males and females combined) statistically controlling for sex and sex-CR interaction. These conditions included: cancer, diabetes, arthritis, diverticulosis and cardiovascular disease. The difference between control and CR is not statistically significant, $P = 0.06$. Old-onset monkeys are not represented.

were detected in control monkeys at an earlier age than in CR monkeys, the incident curves were not significantly different ($P = 0.06$) (Fig. 3b).

Considering that these two projects maintain high quality veterinary support in comparable experimental settings, study the same species of primates, and test the same intervention, what could account for the differences in survival outcome?

A notable difference between the two studies is diet composition. The NIA-1-87 formulation (Labdiet, PMI Nutrition International) has a natural ingredient base whereas WNPRC diet is purified (Harlan Teklad). Although natural ingredient diets risk having some variation between batches, they contain components that may have an impact on health such as phytochemicals, ultra-trace minerals and other unidentified elements¹⁷. In purified diets, each ingredient supplies a specific nutrient and each required mineral and vitamin is added as a separate component. Nutrient sources were also different. Protein was derived from wheat, corn, soybean, fish and alfalfa meal for the NIA diet, whereas the WNPRC diet protein source was lactalbumin. The NIA diet also contained flavonoids, known for their antioxidant activity, and fat from soy oil and the oils from the other natural ingredients (that is, corn, wheat and fish). Fish meal contains approximately 8–12% fat and is rich in omega-3 fatty acids. The WNPRC study dietary fat was derived from corn oil. Carbohydrate content was also notably different; although both diets had 57–61% carbohydrate by weight, the NIA study diet was comprised primarily of ground wheat and corn, whereas the WNPRC study diet contained corn starch and sucrose. Indeed, the WNPRC diet was 28.5% sucrose, whereas the NIA study diet was only 3.9% sucrose. This latter point may be particularly important as a diet high in sucrose may contribute to the incidence of type II diabetes^{18,19}.

The NIA and WNPRC studies also approached vitamin and mineral supplementation differently. The NIA study used one diet for both CR and control monkeys, which was supplemented with an additional 40% of the daily-recommended allowance to insure adequate nutrition for the CR monkeys. Thus, the NIA diet formulation supersupplemented the control monkeys. The WNPRC study fed two different diets and only the CR monkeys were supplemented.

Another important difference in study design was that the NIA study control monkeys were not truly fed *ad libitum*, unlike the WNPRC study. The regulated portioning of food for the NIA control monkeys may be a slight restriction and thus, largely prevented obesity. It has been reported that 10% CR increased lifespan in rats compared to *ad libitum*, even more than 25 and 40% CR²⁰. The NIA control monkeys may experience survival benefits from this slight restriction.

Calorie restriction effectively lowered body weight in the NIA and WNPRC monkeys (Supplementary Table 2)^{13,21}. However, WNPRC monkeys generally weighed more than corresponding NIA monkeys. For example, at 17 years of age, WNPRC males weighed approximately 12% more than corresponding NIA males and the difference was approximately 18% for the females. Thus, the NIA monkeys may be in an optimal weight range.

NIA monkeys originated from both China and India, and have greater genetic diversity compared to the strictly Indian colony at WNPRC. In rodent studies, genetic differences have affected the survival outcome in CR studies, even shortening it in some recombinant inbred mouse strains²². In genetically heterogeneous wild-caught mice, although hormonal and weight loss effects were consistent with CR, there was no overall mean longevity effect²³. It is apparent that the effect of CR is not straightforward, and genetic differences may have a larger role than has been considered to date. A final analysis which includes all monkeys and controls for genetic origin can address this confounding variable.

Lastly, as in rodent studies²⁴, the age of onset of the CR regimen for the two studies could certainly impact survival outcome as it has other measures. CR initiated in the youngest male monkeys delayed maturation²⁵ and slowed skeletal growth²⁶. Additionally, only the immune response of the adolescent males was improved by CR¹⁵.

In the first randomized trial in humans, 6 months of CR improved several biomarkers of ageing and improved cardiovascular health, suggesting a reduction in risk of age-related disease. However, a lifespan study in humans is improbable²⁷. Even a self-imposed CR regimen in lean individuals improved several metabolic, inflammatory and cardiovascular measures²⁸. Current findings show that in nonhuman and human primates, CR evokes very similar metabolic, hormonal and physiological changes that are linked to longevity in CR rodents²⁸. It will be valuable to continue to compare findings from ongoing monkey CR studies to dissect the mechanisms behind the improvement in health that occurred with and without significant effects on survival.

METHODS SUMMARY

Approval. This study was approved by the Animal Care and Use Committee of the NIA, NIH and was conducted in an AALAC-accredited facility.

Diagnostics. In live animals, diagnostic evaluations were made on the basis of clinical presentation. Radiographs confirmed conditions of osteoarthritis; endoscopic evaluation of diverticulosis revealed hernia-like outpouching in the mucosa of the descending colon with trapped faecal material in the diverticula; diabetes was confirmed by consistent elevated fasting glucose and glucose response during an intravenous glucose tolerance test; surgical biopsy or removal of tumours confirmed neoplasia. Cardiovascular abnormalities such as myofibre loss and fibrosis were diagnosed at necropsy as well as death due to acute congestive heart failure.

Blood sampling. For longitudinal measures, blood samples were obtained under ketamine (7–10 mg kg⁻¹, intramuscular) or Telazol (3.5 mg kg⁻¹, intramuscular) anaesthesia following an overnight fast. Serum samples were stored at -80 °C until analysed. Plasma free isoprostane samples were collected in 2005 and measured according to the description in ref. 29.

Full Methods and any associated references are available in the online version of the paper.

Received 27 October 2011; accepted 23 July 2012.

Published online 29 August 2012.

- Branch-Mays, G. L. *et al.* The effects of a calorie-reduced diet on periodontal inflammation and disease in a non-human primate model. *J. Periodontol.* **79**, 1184–1191 (2008).
- Messaoudi, I. *et al.* Delay of T cell senescence by caloric restriction in aged long-lived nonhuman primates. *Proc. Natl Acad. Sci. USA* **103**, 19448–19453 (2006).
- Kastman, E. K. *et al.* A calorie-restricted diet decreases brain iron accumulation and preserves motor performance in old rhesus monkeys. *J. Neurosci.* **30**, 7940–7947 (2010).
- Colman, R. J., Beasley, T. M., Allison, D. B. & Weindruch, R. Attenuation of sarcopenia by dietary restriction in rhesus monkeys. *J. Gerontol. A* **63**, 556–559 (2008).
- Colman, R. J. *et al.* Caloric restriction delays disease onset and mortality in rhesus monkeys. *Science* **325**, 201–204 (2009).
- Bodkin, N. L., Alexander, T. M., Ortmeier, H. K., Johnson, E. & Hansen, B. C. Mortality and morbidity in laboratory-maintained Rhesus monkeys and effects of long-term dietary restriction. *J. Gerontol. A* **58**, B212–B219 (2003).
- Forster, M. J., Morris, P. & Sohal, R. S. Genotype and age influence the effect of caloric intake on mortality in mice. *FASEB J.* **17**, 690–692 (2003).
- Murtagh-Mark, C. M., Reiser, K. M., Harris, R. Jr & McDonald, R. B. Source of dietary carbohydrate affects life span of Fischer 344 rats independent of caloric restriction. *J. Gerontol. A* **50A**, B148–B154 (1995).
- Swindell, W. R. Dietary restriction in rats and mice: a meta-analysis and review of the evidence for genotype-dependent effects on lifespan. *Ageing Res. Rev.* **11**, 254–270 (2012).
- Messaoudi, I. *et al.* in *Calorie Restriction, Aging, and Longevity* (eds Everitt, A. V., Rattan, S., Le Couteur, D. & de Cabo, R.) 55–78 (Springer, 2010).
- Roth, G. S. *et al.* Aging in rhesus monkeys: relevance to human health interventions. *Science* **305**, 1423–1426 (2004).
- Ingram, D. K. *et al.* Dietary restriction and aging: the initiation of a primate study. *J. Gerontol.* **45**, B148–B163 (1990).
- Mattison, J. A. *et al.* Age-related decline in caloric intake and motivation for food in rhesus monkeys. *Neurobiol. Aging* **26**, 1117–1127 (2005).
- Colman, R. J. & Kemnitz, J. W. in *Methods in Aging Research* (ed. Yu, B. P.) 249–267 (CRC, 1998).
- Messaoudi, I. *et al.* Optimal window of caloric restriction onset limits its beneficial impact on T-cell senescence in primates. *Aging Cell* **7**, 908–919 (2008).
- Allison, P. D. *Survival Analysis Using SAS: A Practical Guide* (SAS Institute, 1995).
- Nadon, N. L. Exploiting the rodent model for studies on the pharmacology of lifespan extension. *Aging Cell* **5**, 9–15 (2006).
- Lomba, A. *et al.* A high-sucrose isocaloric pair-fed model induces obesity and impairs NDUFB6 gene function in rat adipose tissue. *J. Nutrigenet. Nutrigenomics* **2**, 267–272 (2009).
- Roncal-Jimenez, C. A. *et al.* Sucrose induces fatty liver and pancreatic inflammation in male breeder rats independent of excess energy intake. *Metabolism* **60**, 1259–1270 (2011).
- Duffy, P. H. *et al.* The effects of different levels of dietary restriction on aging and survival in the Sprague-Dawley rat: implications for chronic studies. *Aging (Milano)* **13**, 263–272 (2011).
- Raman, A. *et al.* Influences of calorie restriction and age on energy expenditure in the rhesus monkey. *Am. J. Physiol. Endocrinol. Metab.* **292**, E101–E106 (2007).
- Liao, C. Y., Rikke, B. A., Johnson, T. E., Diaz, V. & Nelson, J. F. Genetic variation in the murine lifespan response to dietary restriction: from life extension to life shortening. *Aging Cell* **9**, 92–95 (2010).
- Harper, J. M., Leathers, C. W. & Austad, S. N. Does caloric restriction extend life in wild mice? *Aging Cell* **5**, 441–449 (2006).
- Speakman, J. R. & Hambly, C. Starving for life: what animal studies can and cannot tell us about the use of caloric restriction to prolong human lifespan. *J. Nutr.* **137**, 1078–1086 (2007).
- Roth, G. S. *et al.* Age-related changes in androgen levels of rhesus monkeys subjected to diet restriction. *Endocr. J.* **1**, 227–234 (1993).
- Lane, M. A. *et al.* Aging and food restriction alter some indices of bone metabolism in male rhesus monkeys (*Macaca mulatta*). *J. Nutr.* **125**, 1600–1610 (1995).
- Redman, L. M. & Ravussin, E. Caloric restriction in humans: impact on physiological, psychological, and behavioral outcomes. *Antioxid. Redox Signal.* **14**, 275–287 (2011).
- Omodei, D. & Fontana, L. Calorie restriction and prevention of age-associated chronic disease. *FEBS Lett.* **585**, 1537–1542 (2011).
- Ward, W. F. *et al.* Effects of age and caloric restriction on lipid peroxidation: measurement of oxidative stress by F2-isoprostane levels. *J. Gerontol. A* **60**, 847–851 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the animal care staff and technicians, both past and present, especially J. Travis and M. Szarowicz; K. Vaughan for her editorial help; and the many collaborators that have contributed to this project. This research was supported by the Intramural Research Program of the NIH, National Institute on Aging.

Author Contributions G.S.R. and D.K.I. jointly conceived the original study and implemented it. J.A.M., R.d.C., D.K.I. and G.S.R. designed experiments, analysed and discussed data. J.A.M., R.d.C. and D.K.I. wrote the paper. T.M.B. and D.B.A. conducted statistical analyses and consultation. E.M.T., A.M.H. and J.E.Y. provided many years of technical support, data collection and supervision. R.L.H. provided veterinary support. D.L.L. assisted with data interpretation, discussion and paper edits. M.B. performed pathology assessments. D.B. assisted with initial diet formulation and all diet analyses and comparisons. W.F.W. and W.Q. designed and performed the isoprostane assays.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com. Correspondence and requests for materials should be addressed to J.A.M. (mattisonj@mail.nih.gov), R.d.C. (deCaboRa@grc.nia.nih.gov) or D.K.I. (Donald.Ingram@pbrc.edu).

METHODS

Animals. With the exception of six old-onset males, all monkeys had known birthdates. Estimated ages were assigned to these six based on dental archives and historical records. No monkey had been used in invasive experiments before procurement. After procurement, monkeys were initiated on the study after required quarantine. Food intake was considered *ad libitum* during this time. Husbandry has been described previously¹³. NIA monkeys were fed a natural ingredient diet containing 56.9% carbohydrate, 17.3% protein and 5% fat.

Statistical methods. A Fisher's exact test was used to compare the incidence of neoplasia in the young-onset cohort. Analyses of age-associated diseases and mortality included all animals with known diagnoses or cause of death before 1 December 2011.

Twenty of the 26 adult-onset females were obtained from a military research facility, and 19 of these monkeys developed severe and rapidly progressing endometriosis. The twentieth monkey of this group died at the age of 12 years from renal necrosis. It seemed apparent that this cohort was differentially affected in terms of long-term health, and thus, an indicator variable that designated the source of this monkey group as 'Aberdeen' was created and was included in most analyses to control statistically for the effects of these animals on the outcomes of interest.

To determine the effect of CR on the onset of age-associated diseases (morbidity) and mortality, a Cox proportional hazard¹⁶ regressions with sex and caloric restriction (CR), a sex–CR interaction term, and a covariate to adjust for whether the animal was obtained from the Aberdeen site as predictors were used to estimate the survival and hazard functions. The proportional hazards (PH) assumption was tested by fitting a non-PH Cox regression with a CR–time interaction, which was not significant for either analysis, and thus, PH models were considered valid. Animals that died of non-age-related causes (for example, death from anaesthesia, gastrointestinal bloat) were censored in both the mortality and morbidity analyses. Their age at death was used as the time variable in the Cox regressions. For the morbidity analysis, the age at which the animal experienced its first age-related diagnosis was used as the time variable in the Cox regression. Animals that received a non-age-related diagnosis were censored and their current age was used as the time variable. Animals that died of an age-related cause without ever receiving an age-related diagnosis were not censored. The designation of 'age-related' was based on the same rationale and list of conditions as reported by WNPRC. Death was considered as their first age-related diagnosis and their age at death was used as the time variable. All analyses were performed in SAS PROC PHREG and likelihood ratio tests were computed to assess statistical significance.

A linear mixed model³⁰ approach was used to estimate longitudinal trends in the data while accounting for the dependency in the data due to multiple observations per subject. SAS PROC MIXED was used to estimate the trends and group differences among the repeatedly measured outcomes (for example, body weight, glucose, cholesterol and triglycerides across the years of measurement). The effects of CR on overall outcome levels and differences in longitudinal trends were tested by including diet main effect and diet–year interaction terms in the model. Male and female monkeys were analysed together and sex main effects and sex–diet and

–year interactions were also included in the models. The young- and old-onset groups were analysed separately. Age at the first measurement (that is, starting age) was used as a covariate to control for differences in age among the animals within a given year of measurement and a lag-1 autoregressive process over time was assumed. For the young-onset group a covariate to adjust for whether they were obtained from the Aberdeen site was added. Outliers were screened and removed. Specifically, a few young animals had glucose levels substantially above 200 mg dl^{−1}. Also one old control male that was eventually diagnosed with diabetes had extremely high triglyceride levels ranging from 342 to 1,314 mg dl^{−1} and these values influenced the significance of some effects. Briefly, the linear mixed model approach estimates a growth trajectory for each individual animal (for example, individual change in weight over time), adjusting for covariates. Then a weighted composite of these individual trajectories is computed to show the average trend over the age of the animals in a particular group (for example, average weight of animals at varying ages for CR-M). The weights for these composites are based on the number of observations each animal contributes to the data. For example, animals that live longer will contribute more data, and therefore will get larger weights. To smooth the trends for plotting graphs, the predicted values from each individual trajectory was averaged, and loess trend lines were constructed.

Competing risk. The analyses in this paper as well as in ref. 5 distinguished between age-related and all-cause mortality. To address the issue that the non-age-related deaths are associated to CR, a competing risks Cox proportional hazard regression models¹⁶ were conducted separately for the young-onset group (9 control and 13 CR non-age-related deaths) and old-onset (2 control and 3 CR non-age-related deaths). Briefly, a competing risks model treats the events as if age-related and non-age-related deaths are mutually exclusive and compared to neither event occurring (that is, animals still alive are censored). These events have competing risks in that if an animal dies from a non-age-related cause they are no longer at-risk for an age-related death (and vice versa). For the old-onset animals, age at start of the experiment was not significantly related to non-age- ($P = 0.188$) or age-related deaths ($P = 0.269$). CR was not significantly related to non-age- ($P = 0.260$) or age-related deaths ($P = 0.490$). Also, sex was not significantly related to non-age- ($P = 0.991$) or age-related deaths ($P = 0.053$); however, this association of sex with age-related mortality is of marginal significance, which is consistent with the trend for males for have higher survival curves (see Figs 1a and 2a, b). For the young-onset animals, age at start of the experiment was not significantly related to non-age- ($P = 0.604$) or age-related deaths ($P = 0.653$). Sex was not significantly related to non-age- ($P = 0.790$) or age-related deaths ($P = 0.480$). CR was not significantly related to non-age- ($P = 0.147$) or age-related deaths ($P = 0.975$). Also, the origin (Aberdeen) of the animal was not significantly related to age-related deaths ($P = 0.513$), and the relationship to non-age-related deaths was not statistically significant. This marginal P value ($P = 0.0889$) could suggest that origin may be a confounding factor.

30. Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. & Schabenberger, O. SAS for Mixed Models (SAS Institute Inc., 2006).

Increased proteasome activity in human embryonic stem cells is regulated by PSMD11

David Vilchez¹, Leah Boyer², Ianessa Morantte¹, Margaret Lutz³, Carsten Merkwirth¹, Derek Joyce¹, Brian Spencer⁴, Lesley Page⁵, Eliezer Masliah⁴, W. Travis Berggren³, Fred H. Gage² & Andrew Dillin¹

Embryonic stem cells can replicate continuously in the absence of senescence and, therefore, are immortal in culture^{1,2}. Although genome stability is essential for the survival of stem cells, proteome stability may have an equally important role in stem-cell identity and function. Furthermore, with the asymmetric divisions invoked by stem cells, the passage of damaged proteins to daughter cells could potentially destroy the resulting lineage of cells. Therefore, a firm understanding of how stem cells maintain their proteome is of central importance. Here we show that human embryonic stem cells (hESCs) exhibit high proteasome activity that is correlated with increased levels of the 19S proteasome subunit PSMD11 (known as RPN-6 in *Caenorhabditis elegans*)^{3–5} and a corresponding increased assembly of the 26S/30S proteasome. Ectopic expression of PSMD11 is sufficient to increase proteasome assembly and activity. FOXO4, an insulin/insulin-like growth factor-I (IGF-I) responsive transcription factor associated with long lifespan in invertebrates^{6,7}, regulates proteasome activity by modulating the expression of PSMD11 in hESCs. Proteasome inhibition in hESCs affects the expression of pluripotency markers and the levels of specific markers of the distinct germ layers. Our results suggest a new regulation of proteostasis in hESCs that links longevity and stress resistance in invertebrates to hESC function and identity.

Embryonic stem cells are unique among all stem-cell populations examined in that they do not seem to undergo replicative senescence^{1,2}. Because proteostasis is crucial for maintaining proper cell function^{8,9}, hESCs could provide a new model to define proteostasis regulation and its demise in ageing. Central to proteostasis is the ubiquitin proteasome system and we examined whether proteasome activity changes as hESCs differentiate into several cell lineages. To evaluate differences in the 26S/30S proteasome activity, we monitored the degradation of specific fluorogenic peptide substrates¹⁰. We differentiated H9 hESCs into neural progenitor cells (NPCs) and observed a marked decrease in the chymotrypsin-like proteasome activity (Fig. 1a). Moreover, when NPCs were differentiated into neurons, we detected a further decrease in proteasome activity during the differentiation process that was observable after 2 weeks (Fig. 1a and Supplementary Fig. 1). Consistent with enhanced proteasome activity in hESCs, we found increased levels of polyubiquitinated proteins in differentiated cells compared with hESCs (Fig. 1b). Because hESCs are known to vary in their characteristics despite unlimited capacity of self-renewal¹¹, we differentiated a distinct hESC line, HUES-6 cells, and obtained similar results (Supplementary Figs 1 and 2). Proteasome inhibitors blocked activity from extracts of hESCs, NPCs and neurons (Supplementary Fig. 3), indicating that the increased peptidase activity was indeed due to the proteasome. In addition, the other two activities of the proteasome, the caspase-like and trypsin-like activities, were also increased in hESCs (Supplementary Fig. 4). Proteasome activity did not differ depending on the passage number (Supplementary Fig. 5).

The decrease in proteasome activity was not a specific phenomenon associated with the neural lineage as differentiation into either trophoblasts or fibroblasts induced a similar decrease (Fig. 1c, d).

Notably, hESCs lost their high proteasome activity in a continuous progressive manner during the differentiation process (Fig. 1c). Moreover, we examined other cell lines extracted from human tissues, such as astrocytes, BJ fibroblasts or immortalized HEK293T cells, and found that these cells also had lower proteasome activity compared with hESCs (Supplementary Fig. 6). We tested whether high proteasome activity in hESCs was associated with increased proliferation and found

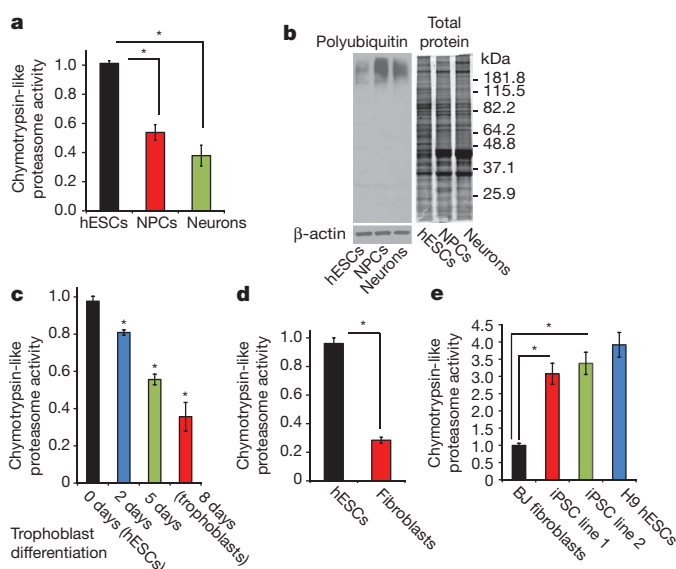


Figure 1 | Increased proteasome activity in hESCs. **a**, Chymotrypsin-like (Z-Gly-Gly-Leu-AMC) proteasome activity (relative slope to H9 hESCs) is shown. hESCs (n = 11), NPCs (n = 13), neurons (n = 10); *P < 0.00001. **b**, Representative immunoblot of polyubiquitinated protein levels. β-actin is the loading control. Total protein was visualized by Coomassie staining. **c**, Chymotrypsin-like proteasome activity (relative slope to H9 hESCs) is shown. hESCs (n = 6), 2 days of differentiation into trophoblasts (n = 6), 5 days (n = 6), 8 days (n = 7). hESCs versus 2 days, *P < 0.001; hESCs versus 5 days, *P = 7.7 × 10⁻⁷; hESCs versus 8 days, *P < 0.0001. **d**, Chymotrypsin-like proteasome activity (relative slope to H9 hESCs) is shown. hESCs (n = 3), fibroblasts (n = 3); *P < 0.001. **e**, Chymotrypsin-like proteasome activity (relative slope to BJ fibroblasts) is shown. BJ fibroblasts (n = 10), iPSC line 1 (n = 10), iPSC line 2 (n = 9), H9 hESCs (n = 5). iPSC lines have increased proteasome activity compared with fibroblasts (*P < 0.0005), and no significant differences compared with H9 hESCs (iPSC line 1 versus hESCs, P = 0.11; iPSC line 2 versus hESCs, P = 0.29). In a, c, d and e data represent the mean ± s.e.m. All statistical comparisons were made by Student's t-test for unpaired samples.

¹Howard Hughes Medical Institute, Glenn Center for Aging Research, Molecular and Cell Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ²Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ³Stem Cell Core, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. ⁴Department of Neurosciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. ⁵Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

that hESCs and HEK293T cells had nearly identical proliferation rates, yet hESCs had higher proteasome activity (Supplementary Fig. 7). Induced pluripotent stem cells (iPSCs) can be derived from adult somatic cells by forced expression of exogenous factors that promote cell reprogramming^{12–14}. iPSC lines are similar to embryonic stem cells in many respects, such as their gene expression patterns, proteome profile and potential for differentiation^{13,15}. However, the full extent of their similarity to embryonic stem cells is still being assessed¹⁶. We analysed two iPSC lines that had been carefully validated to ensure similar

gene expression profiles, growth characteristics and developmental potentials to hESCs¹⁷. We discovered that these iPSC lines derived from BJ fibroblasts display increased proteasome activity similar to hESCs (Fig. 1e), indicating that proteasomal activity can indeed be reprogrammed.

The 26S/30S proteasome consists of a 20S core structure containing the proteolytic active sites and 19S cap structures that impart regulation on the activity of the holo-complex (26S, single and 30S, double capped)¹⁸. Although 20S particles can exist in a free form, 20S particles

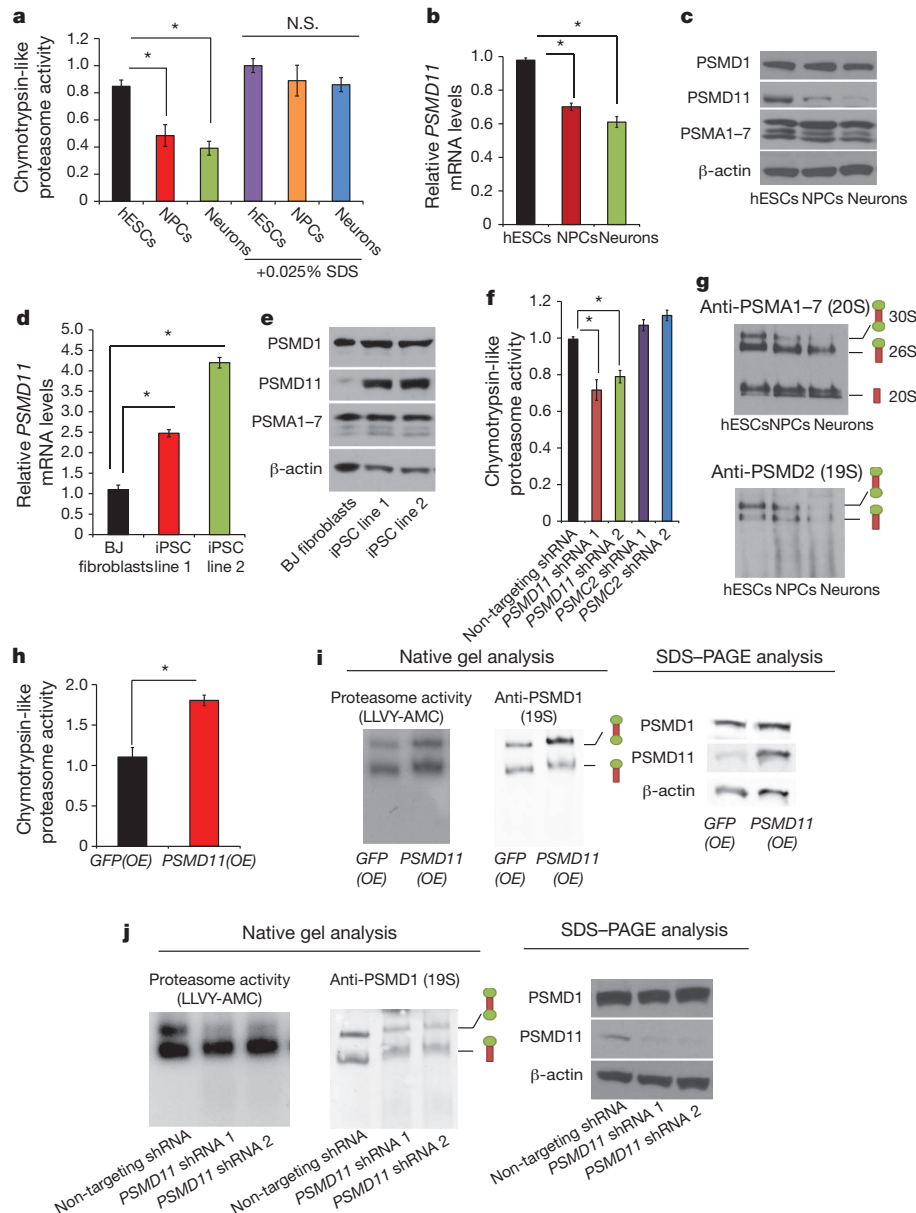


Figure 2 | Increased proteasome assembly in hESCs depends on PSMD11 expression. **a**, Chymotrypsin-like proteasome activity (relative slope to H9 hESCs + 0.025% SDS) is shown. $n = 4–6$; * $P < 0.01$. No significant differences (N.S.) were found among the different cells when SDS was added (hESCs + SDS versus NPCs + SDS, $P = 0.25$; hESCs + SDS versus neurons + SDS, $P = 0.09$). **b**, Relative PSMD11 expression compared with H9 hESCs. hESCs ($n = 10$), NPCs ($n = 6$), neurons ($n = 8$); * $P < 0.00001$. **c**, **e**, Western blot analysis with antibodies against PSMA1–7, PSMD11 and PSMD1. β -actin is the loading control. **d**, Relative PSMD11 expression compared with BJ fibroblasts. BJ fibroblasts ($n = 10$), iPSC line 1 ($n = 6$), iPSC line 2 ($n = 6$); * $P < 0.00001$. **f**, Proteasome activity (relative slope to lentivirus non-targeting shRNA) is shown. Non-targeting shRNA ($n = 10$), PSMD11 shRNA 1 ($n = 8$), PSMD11

shRNA 2 ($n = 6$), PSMC2 shRNA 1 ($n = 6$), PSMC2 shRNA 2 ($n = 3$); * $P < 0.01$. **g**, Native gel electrophoresis followed by Western blot with PSMA1–7 (20S subunit) or PSMD2 (19S subunit) antibodies. **h**, Chymotrypsin-like proteasome activity (relative slope to green fluorescent protein (GFP) overexpressing (OE) HEK293T cells) is shown. GFP(OE) ($n = 4$), PSMD11(OE) ($n = 5$); * $P < 0.005$. **i**, **j**, Native gel electrophoresis of HEK293T cell extracts followed by proteasome activity assay with chymotrypsin-like activity substrate Leu-Leu-Val-Tyr-AMC and immunoblotting with PSMD1 (19S subunit) antibody are shown. Extracts were resolved by SDS-PAGE and immunoblotting for analysis of PSMD11 levels and β -actin loading control. In **a**, **b**, **d**, **f** and **h**, data represent the mean \pm s.e.m. All statistical comparisons were made by Student's t -test for unpaired samples.

in their most physiological form are inactive, unable to degrade denatured proteins or cleave peptides¹⁰. The 19S regulatory subunit is responsible for stimulating the 20S proteasome to degrade proteins, because ATPases of the regulatory particle open the 20S core, allowing substrates access to proteolytic active sites¹⁹. SDS treatment of cell extracts, a condition that activates 20S particles by allowing gate opening²⁰, resulted in equivalent activities among all cell types (Fig. 2a and Supplementary Fig. 8). This result suggests that all cell types have an equal number of 20S particles, but hESCs have increased levels of active 26S/30S proteasomes. We examined the expression of the different 19S proteasome subunits and observed that PSMD11 was the only 19S subunit to decrease as hESCs differentiated (Fig. 2b, c and Supplementary Figs 9–12). Consistent with the hESCs results, we observed increased PSMD11 levels in iPSCs (Fig. 2d, e). Accordingly, decreased expression of PSMD11 in hESCs (Supplementary Table 1) reduced proteasome activity (Fig. 2f), demonstrating that the increased levels of this subunit in hESCs are crucial for increased proteasome activity. PSMD11 has a vital role in stabilizing the otherwise weak interaction between the 20S core and the 19S cap⁴, suggesting that hESCs might have more assembled proteasomes. We detected more 30S particles in hESCs compared with NPCs and neurons. Furthermore, as more 20S subunits are assembled into 30S particles, less free 20S is found in hESCs (Fig. 2g). Ectopic expression of PSMD11 was sufficient to increase proteasome activity and assembly in cells with relatively low proteasome activity (Fig. 2h, i). Moreover, knockdown of PSMD11 resulted in fewer assembled proteasomes (Fig. 2j).

The levels of RPN-6, the *C. elegans* orthologue of PSMD11, are increased in the long-lived *glp-1* mutant. In this mutant, increased proteasome activity, *rpn-6* expression and longevity are modulated by the forkhead box O (FOXO) transcription factor DAF-16. To examine whether FOXO transcription factors regulate proteasome activity in hESCs, we reduced expression of the closest human

daf-16 orthologues FOXO1, FOXO3 and FOXO4 (Supplementary Fig. 13 and Supplementary Table 2). Notably, we found that FOXO4 was crucial to modulate proteasome activity in hESCs, whereas FOXO1 and FOXO3, as well as a HSF1, had little or no effect on proteasome activity (Fig. 3a, b and Supplementary Fig. 14). As hESCs differentiated into neural cells, trophoblasts or fibroblasts, there was a corresponding decrease in FOXO4 expression (Supplementary Figs 15–17). Accordingly, this decrease in FOXO4 expression is reprogrammed from somatic cells to iPSCs (Supplementary Fig. 15). FOXO1 had a similar expression pattern to FOXO4 in H9 but not in HUES-6 hESCs (Supplementary Figs 15 and 16). Furthermore, reduction of FOXO4 affected proteasome activity in the multipotent NPCs, which retain partial stem-cell character, but did not affect proteasome activity in differentiated neurons (Supplementary Fig. 18 and Supplementary Table 3). These results raised the question as to whether FOXO4 regulation of proteasome activity could be a general mechanism found in dividing cells. However, we found that FOXO4 was not required for proteasome activity regulation in BJ fibroblasts or HEK293T but rather seems to be specific to hESCs (Supplementary Fig. 19). FOXO4 transcriptional activity is inhibited by phosphorylation on Thr 32, Ser 197 and Ser 262 sites, and once dephosphorylated, it translocates to the nucleus and induces target gene expression^{21,22}. Expression of a constitutively active FOXO4 triple alanine mutant (FOXO4-AAA), but not wild-type FOXO4, resulted in further upregulation of proteasome activity in hESCs (Fig. 3c, Supplementary Fig. 20 and Supplementary Table 4a, b). In addition, ectopic expression of FOXO4-AAA in FOXO4 short hairpin RNA (shRNA) cells partially restored proteasome activity of these cells (Fig. 3d and Supplementary Table 4c). We found that loss of FOXO4 resulted in reduced expression of PSMD11 in hESCs and in the multipotent NPCs, but did not affect PSMD11 in differentiated neurons (Fig. 3e, f and Supplementary Fig. 21). Knockdown of FOXO1, FOXO3 or HSF1 did not affect

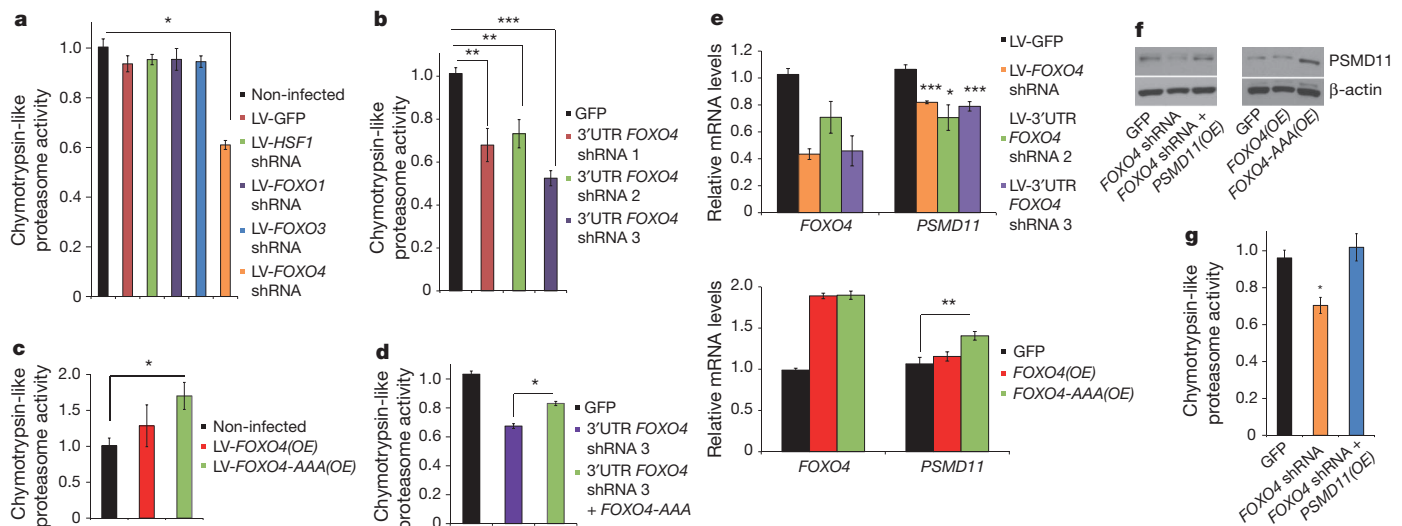


Figure 3 | FOXO4 regulates proteasome activity in hESCs.

a, Chymotrypsin-like proteasome activity in H9 hESCs transiently infected with lentiviruses (LV; relative slope to non-infected cells) is shown. $n = 19$; $*P < 0.00001$. **b**, Chymotrypsin-like proteasome activity of stable-infected hESCs (relative slope to GFP cells) is shown. GFP ($n = 7$), FOXO4 shRNA 1 ($n = 6$), FOXO4 shRNA 2 ($n = 3$), FOXO4 shRNA 3 ($n = 6$). GFP versus FOXO4 shRNA 1, $**P < 0.01$; GFP versus FOXO4 shRNA 2, $**P < 0.01$; GFP versus FOXO4 shRNA 3, $***P = 4.5 \times 10^{-8}$. All shRNAs target the 3' untranslated region (UTR) of the human genes. **c**, Proteasome activity of transiently infected hESCs (relative slope to non-infected H9 hESCs) is shown. $n = 7$; non-infected cells versus FOXO4(OE) cells, $*P = 0.41$; non-infected cells versus FOXO4-AAA(OE) cells, $*P < 0.05$. **d**, Proteasome activity (relative slope to GFP hESCs) is shown. $n = 4$; FOXO4 shRNA 3 versus FOXO4 shRNA 3 + FOXO4-AAA, $*P < 0.01$. **e**, Top, knockdown of FOXO4 decreases expression of PSMD11 in transiently infected H9 hESCs (GFP versus FOXO4 shRNA, $***P < 0.001$; GFP versus FOXO4 shRNA 2, $*P < 0.05$; GFP versus FOXO4 shRNA 3, $***P < 0.001$). GFP ($n = 15$), FOXO4 shRNA ($n = 19$), FOXO4 shRNA 2 ($n = 4$), FOXO4 shRNA 3 ($n = 5$). Bottom, stable overexpression of the FOXO4-AAA mutant increases PSMD11 expression in H9 hESCs (GFP versus FOXO4(OE), $P = 0.69$; GFP versus FOXO4-AAA(OE), $***P < 0.01$). GFP ($n = 7$), FOXO4(OE) ($n = 8$), FOXO4-AAA(OE) ($n = 7$). **f**, Western blot analysis of PSMD11 levels. β -actin is the loading control. **g**, Proteasome activity (relative slope to GFP H9 hESCs) is shown. $n = 4$; GFP versus FOXO4 shRNA, $*P < 0.01$; GFP versus FOXO4 shRNA + PSMD11(OE), $P = 0.50$. Data in **a–e** and **g** represent the mean \pm s.e.m. Statistical comparisons were made by Student's *t*-test for unpaired samples.

3 + FOXO4-AAA, $*P < 0.01$. **e**, Top, knockdown of FOXO4 decreases expression of PSMD11 in transiently infected H9 hESCs (GFP versus FOXO4 shRNA, $***P < 0.001$; GFP versus FOXO4 shRNA 2, $*P < 0.05$; GFP versus FOXO4 shRNA 3, $***P < 0.001$). GFP ($n = 15$), FOXO4 shRNA ($n = 19$), FOXO4 shRNA 2 ($n = 4$), FOXO4 shRNA 3 ($n = 5$). Bottom, stable overexpression of the FOXO4-AAA mutant increases PSMD11 expression in H9 hESCs (GFP versus FOXO4(OE), $P = 0.69$; GFP versus FOXO4-AAA(OE), $***P < 0.01$). GFP ($n = 7$), FOXO4(OE) ($n = 8$), FOXO4-AAA(OE) ($n = 7$). **f**, Western blot analysis of PSMD11 levels. β -actin is the loading control. **g**, Proteasome activity (relative slope to GFP H9 hESCs) is shown. $n = 4$; GFP versus FOXO4 shRNA, $*P < 0.01$; GFP versus FOXO4 shRNA + PSMD11(OE), $P = 0.50$. Data in **a–e** and **g** represent the mean \pm s.e.m. Statistical comparisons were made by Student's *t*-test for unpaired samples.

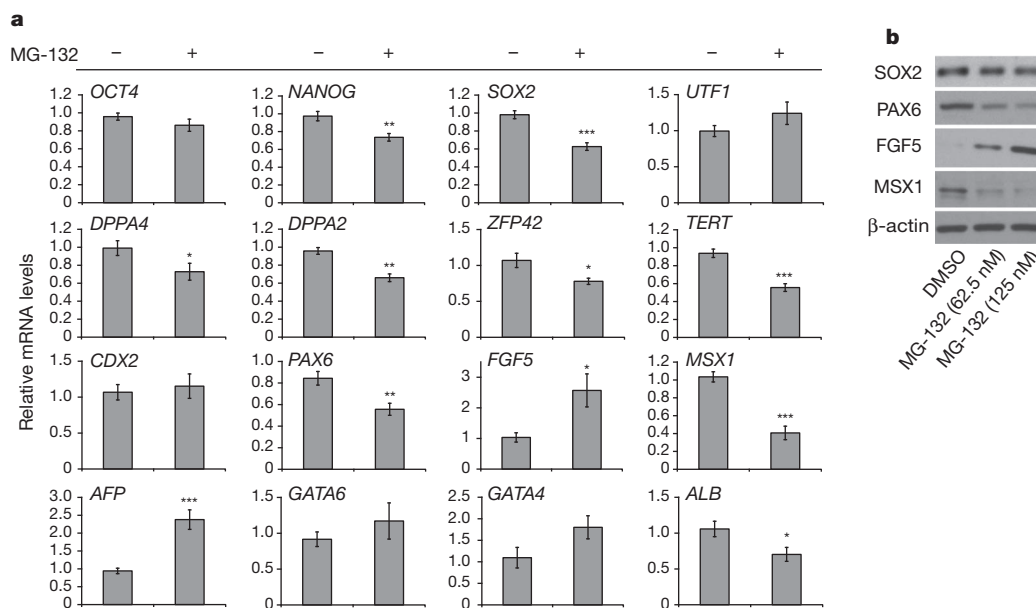


Figure 4 | Acute proteasome inhibition affects pluripotency of hESCs.
a, Real-time PCR analysis of pluripotency (*OCT4* (also known as *POU5F1*), *NANOG*, *SOX2*, *UTF1*, *DPPA4*, *DPPA2*, *ZFP42* and *TERT*), trophectodermal (*CDX2*), ectodermal (*PAX6*, *FGF5*), mesodermal (*MSX1*) and endodermal (*AFP*, *GATA6*, *GATA4* and *ALB*) germ-layer markers. The graphs (relative expression to H9 hESCs plus dimethylsulphoxide (DMSO)) represent the

mean \pm s.e.m., $n = 12$. Proteasome inhibition (62.5 nM MG-132 for 24 h) in hESCs induces a decrease in pluripotency markers and modifies the levels of germ-layer markers. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. **b**, Western blot analysis with antibodies against *SOX2*, *PAX6*, *FGF5* and *MSX1*. β -actin is the loading control. Statistical comparisons were made by Student's *t*-test for unpaired samples.

the expression of PSMD11 in any cell type tested, including hESCs (Supplementary Tables 5–7). Moreover, overexpression of FOXO4-AAA increased PSMD11 levels in hESCs (Fig. 3e, f and Supplementary Table 5). Ectopic expression of PSMD11 in FOXO4 shRNA hESCs rescued their proteasome activity, indicating that expression of PSMD11 by FOXO4 is sufficient to regulate proteasomal function (Fig. 3f, g).

To test the effects of proteasome inhibition on hESC identity, we induced an acute proteasome inhibition in hESCs by using the MG-132 proteasome inhibitor. Notably, hESCs were more sensitive to proteasome inhibition than were NPCs or neurons, and we had to decrease the MG-132 concentration almost 100 times to avoid cell death and detachment of hESCs (data not shown). Notably, low concentrations of MG-132 (62.5 nM) were sufficient to reduce proteasome activity and induce accumulation of polyubiquitinated proteins in hESCs (Supplementary Fig. 22). Even in the absence of differentiation treatment, we already observed that proteasome inhibition resulted in decreased levels of pluripotency markers and modified the levels of markers of the distinct germ and extraembryonic layers, while decreasing the expression of proteins involved in neurogenesis, such as *PAX6* and *MSX1* (Fig. 4).

Collectively, our results establish increased proteasome activity as an intrinsic characteristic of hESC identity. Our findings raise the intriguing question of why these cells need enhanced proteasome activity. One possibility is that hESCs cannot tolerate toxic, misfolded proteins, and increased proteostasis could be required to avoid hESC senescence and maintain an intact proteome for either self-renewal or the generation of an intact cell lineage. Alternatively, high proteasome activity may be tightly linked to other cellular process, such as translation, to ensure future integrity of the proteome. In addition, our results indicate that an orthologue of DAF-16, a transcription factor that regulates both lifespan and resistance to proteotoxic stress in invertebrates, crosses evolutionary boundaries and links hESC identity to invertebrate longevity modulation. It will be of particular interest to identify other genes of the proteostasis network regulated by FOXO4 in hESCs. In conclusion, our findings may trigger new advances in understanding

hESC differentiation or cell reprogramming and open new possibilities for cell therapy by modulation of the proteostasis network.

METHODS SUMMARY

26S proteasome fluorogenic peptidase assays. The *in vitro* assay of 26S proteasome activities was performed as previously described¹⁰. Cells were collected in proteasome activity assay buffer (50 mM Tris-HCl, pH 7.5, 250 mM sucrose, 5 mM MgCl₂, 0.5 mM EDTA, 2 mM ATP and 1 mM dithiothreitol) and lysed by passing ten times through a 27-gauge needle. Lysate was centrifuged at 10,000g for 10 min at 4 °C. Fifteen to twenty-five micrograms of total protein of cell lysates were transferred to a 96-well microtiter plate (BD Falcon) and incubated with fluorogenic substrate. Fluorescence (380 nm excitation, 460 nm emission) was monitored on a microplate fluorometer (Infinite M1000, Tecan) every 5 min for 1 h at 37 °C.

Native gel immunoblotting of the proteasome. HEK293T cells were run on 3.5% native gels prepared in resolving buffer (90 mM Tris base, 90 mM boric acid, 5 mM MgCl₂, 0.5 mM EDTA and 1 mM ATP) with 5 mM ATP, 1 mM dithiothreitol and 3.5% acrylamide from a 40% stock solution of acrylamide and bisacrylamide in a 37.5:1 ratio (Bio-Rad). These were run at 110 V for 3 h at 4 °C. Activity assays were performed by incubating the gels in activity assay buffer for 20 min at 37 °C and developed using a BioRad Gel Doc with ultraviolet illumination. Before transfer, the gels were incubated in transfer buffer (25 mM Tris base and 192 mM glycine) with 1% SDS for 10 min followed by a 10-min incubation in transfer buffer. The protein was transferred to polyvinylidene difluoride (PVDF) membranes at 5 V for 16 h in transfer buffer using an Idea Scientific Genie Blotter. Western blot analysis was performed with anti-PSMD1 (Abcam) and analysed using the Odyssey system (LI-COR). Extracts were also analysed by SDS-PAGE to determine protein expression levels and loading control.

A detailed description of all experimental methods, including hESC culture and differentiation, is provided in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 8 August 2011; accepted 6 August 2012.

- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145–1147 (1998).
- Isono, E., Saito, N., Kamata, N., Saeki, Y. & Toh, E. A. Functional analysis of Rpn6p, a lid component of the 26 S proteasome, using temperature-sensitive *rpn6*

- mutants of the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 6537–6547 (2005).
4. Pathare, G. R. *et al.* The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. *Proc. Natl Acad. Sci. USA* **109**, 149–154 (2012).
 5. Santamaria, P. G., Finley, D., Ballesta, J. P. & Remacha, M. Rpn6p, a proteasome subunit from *Saccharomyces cerevisiae*, is essential for the assembly and activity of the 26 S proteasome. *J. Biol. Chem.* **278**, 6687–6695 (2003).
 6. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A. C. *elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
 7. Tatar, M. *et al.* A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292**, 107–110 (2001).
 8. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916–919 (2008).
 9. Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W. & Balch, W. E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **78**, 959–991 (2009).
 10. Kisselev, A. F. & Goldberg, A. L. Monitoring activity and inhibition of 26S proteasomes with fluorogenic peptide substrates. *Methods Enzymol.* **398**, 364–378 (2005).
 11. Osafune, K. *et al.* Marked differences in differentiation propensity among human embryonic stem cell lines. *Nature Biotechnol.* **26**, 313–315 (2008).
 12. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
 13. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
 14. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
 15. Hanna, J. H., Saha, K. & Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143**, 508–525 (2010).
 16. Panopoulos, A. D., Ruiz, S. & Izpisua Belmonte, J. C. iPSCs: induced back to controversy. *Cell Stem Cell* **8**, 347–348 (2011).
 17. Brennand, K. J. *et al.* Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221–225 (2011).
 18. Finley, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**, 477–513 (2009).
 19. Köhler, A. *et al.* The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Mol. Cell* **7**, 1143–1152 (2001).
 20. Coux, O., Tanaka, K. & Goldberg, A. L. Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.* **65**, 801–847 (1996).
 21. Kops, G. J. P. L. *et al.* Direct control of the Forkhead transcription factor AFX by protein kinase B. *Nature* **398**, 630–634 (1999).
 22. Matsuzaki, H., Ichino, A., Hayashi, T., Yamamoto, T. & Kikkawa, U. Regulation of intracellular localization and transcriptional activity of FOXO4 by protein kinase B through phosphorylation at the motif sites conserved among the FOXO family. *J. Biochem.* **138**, 485–491 (2005).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank W. E. Balch for critical comments on this work. We thank A. Stauffer and V. Modesto for their help with BrdU assays and cell culture, respectively. We thank S. Ruiz for advice on hESC culture and lentiviral infection. This work was supported by the Howard Hughes Medical Institute. D.V. was a recipient of the F.M. Kirby, Inc. Foundation Postdoctoral Scholar Award and Beatriz de Pinós (AGAUR) fellowship. F.H.G. acknowledges the Helmsley Foundation, JPB Foundation, Mathers Foundation, Lookout Fund and California Institute for Regenerative Medicine.

Author Contributions D.V. and A.D. planned and supervised the project. D.V. performed the experiments, data analysis and interpretation. L.B. performed neural differentiation assays and contributed to other assays. I.M. performed biochemistry experiments and contributed to other assays. M.L. performed cell culturing and trophoblast/fibroblast differentiation. C.M. performed biochemistry experiments and contributed to other assays. D.J. performed proteasome assembly experiments. B.S., L.P. and E.M. generated lentiviral constructs. W.T.B. and F.H.G. contributed with their knowledge of stem-cell biology and neural differentiation, and helped to supervise the project. The manuscript was written by D.V. and A.D. and edited by L.B., I.M., C.M., W.T.B. and F.H.G. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D. (dillin@salk.edu).

METHODS

hESCs culture and differentiation. The human H9 (WA09) hESC line was obtained from WiCell Research Institute. The HUES-6 hESC line was obtained from the laboratory of D. Melton. hESC lines were maintained on a mitotically inactive mouse embryonic fibroblast (MEF) feeder layer in hESC medium, DMEM/F12 (Invitrogen) supplemented with 20% knockout serum replacement (Invitrogen), 1 mM L-glutamine, 0.1 mM non-essential amino acids, β -mercaptoethanol and 10 ng ml⁻¹ bFGF (Joint Protein Central). When the co-culturing of hESCs with MEFs was not possible owing to interference with downstream assays, H9 hESCs were also maintained on Matrigel (BD Biosciences) using mTeSR1 (Stem Cell Technologies). When cultured on Matrigel, HUES-6 cells were fed conditioned medium collected from cultured MEFs. hESC colonies were passaged using a solution of collagenase (1 mg ml⁻¹) or dispase (2 mg ml⁻¹), and scraping the colonies with a glass pipette. For our experimental assays, we used H9 hESCs passages 40–45 and HUES-6 hESCs passages 30–35. The human iPSC lines (control BJ-iPSC lines) were derived and characterized as previously reported¹⁷ and cultured as described above for hESCs cells.

Neural differentiation was performed as follows. hESCs grown on inactivated MEFs were fed N2/B27 medium (DMEM/F12-GlutaMAX (Invitrogen), N2 (Invitrogen) B27 without retinoic acid (Invitrogen)) for 2 days before being treated with collagenase type IV (1 mg ml⁻¹ in DMEM/F12) at 37 °C for 1 h. Once colonies lifted off the plate, they were gently washed and transferred to ultra-low attachment plates (Corning). Aggregates (embryoid bodies) were allowed to form and grown in suspension for 1 week in N2/B27 medium with medium changes as needed, roughly every other day. Embryoid bodies were then transferred onto polyornithine (PORN)/laminin-coated plates in N2/B27 medium with 1 μ g ml⁻¹ laminin (Invitrogen), where they were allowed to adhere and develop neural rosettes and projections. After 1 week, colonies were picked for NPC lines. Picked colonies containing rosettes or projections are dissociated with TrypLE (Invitrogen) for 5 min at 37 °C and plated onto PORN/laminin coated plates in NPC medium (DMEM/F12, N2/B27-RA (Invitrogen), 1 μ g ml⁻¹ laminin and 20 ng ml⁻¹ FGF2). The resulting monolayer culture was grown at a high density and split 1:3 every week. For our experimental assays, we used NPCs at passages 10–14.

For neuronal differentiation, NPCs were dissociated with TrypLE (Invitrogen) and plated into neuronal differentiation medium (DMEM/F12, N2/B27-RA (Invitrogen), 1 μ g ml⁻¹ laminin, 20 ng ml⁻¹ BDNF (Peprotech), 20 ng ml⁻¹ GDNF (Peprotech), 1 mM dibutyl-cyclic AMP (Sigma) and 200 nM ascorbic acid (Sigma)) onto PORN/laminin-coated plates. For this study, cells were differentiated in 6-well plates, with approximately 2×10^5 cells per well. Cells were differentiated for 2–3 months, with weekly feeding of neuronal differentiation medium.

Differentiation to fibroblast cells involved the formation of embryoid bodies as described above, but cultured in embryoid body medium (IMDM base medium supplemented with 15% FBS (Atlanta Biologicals), 0.1 mM non-essential amino acids and 1% glutamax (Invitrogen)) and maintained on ultra-low attachment plates with daily medium changes. One week later the floating embryoid bodies were plated on gelatin-coated plates and passaged at confluence three times before use. Alternatively, a non-embryoid body method was used involving the individualization of hESCs using Accutase (Millipore) and plating the cells at a density of 2.5×10^4 cells per cm² in embryoid body medium containing ROCK inhibitor (Stemgent) at 10 μ M. The cells were fed daily with straight embryoid body medium. At confluence, any areas still showing a stem-cell-like morphology were removed by aspiration and then passaged using Accutase. After three passages, the cells present had fibroblast morphology and were confirmed by PCR of lineage-specific markers.

Trophoblast differentiation was performed as described previously using high levels of BMP4 (ref. 23). Keratinocyte differentiation was performed following the protocol established in ref. 24. BJ human fibroblasts (ATCC) were cultured in DMEM (Invitrogen) supplemented with 10% FBS and 0.1 mM non-essential amino acids, and passaged with trypsin. Hippocampal and cerebellar astrocytes are from ScienCell.

Generation of lentiviral vectors. The shRNA-expressing lentiviral vectors were generated by cloning the sequences described in Supplementary Table 8 into the pSIH1-copGFP vector (SBI Biosystems) to generate pLV-siHsf-1, pLV-siFOXO1, pLV-siFOXO3, pLV-siFOXO4, pLV-si3'UTR_1 FOXO4, pLV-si3'UTR_2 FOXO4 and pLV-si3'UTR_3 FOXO4. A control shRNA vector was generated by cloning the sequence 5'-CGTGCCTTGTAGTACTAATCTATTT-3' designed against the sequence of luciferase (SBI Biosystems) into the same vector to generate pLV-siLuc. The GFP-expressing vector was prepared from the third generation self-inactivating lentivirus²⁵. Lentiviruses were packaged by transient transfection in 293T cells²⁵.

Lentivirus (LV)-non targeting shRNA control, LV-shPSMD11_1 (clone number TRCN0000003948), LV-shPSMD11_2 (TRCN0000003950), LV-shPSMC2_1 (TRCN0000007181), LV-shPSMC2_2 (TRCN0000007183) in pLKO.1-puro-CMV-tGFP vector were obtained from Mission shRNA (Sigma). FOXO4-overexpressing lentiviral constructs (FOXO4(OE)) were generated as follows. Flag-FOXO4 construct was obtained from Addgene (plasmid 17549). PCR was performed to generate a product to be cloned into pLVX puro lentiviral plasmid (Clontech) using the XhoI/SmaI sites. Forward primer (with 5' XhoI site for cloning): 5'-CGCGTACTCGAGATGGATCCGGGAATGAGAATTCAGC CACAGAGGCTGCCGCGATCATAGAC-3'; reverse primer (with 3' SmaI site for cloning): 5'-CCGGAACCCGGGTCAGGGATCTGGCTCAAAG-3'. To generate FOXO4-AAA (Thr 32, Ser 197, Ser 262), site-directed mutagenesis of wild-type FOXO4 was performed by using Pfu Turbo. The primers used for site-directed mutagenesis were: Thr32Ala, forward, 5'-GTCCCCGCTCCT GTGCTTGGCCCTTCC-3'; reverse, 5'-GGAAGGGGCCAAGCACAGGAG CGGGGAC-3'. Ser197Ala, forward, 5'-GCAAAGCCCCCGCCGAGAGCC GCAGCCATGGATAGCAGCAG-3'; reverse, 5'-CTGCTGCTATCCATGGT CGGGCTCTGCGCGGGGGGCTTTC-3'. Ser262Ala, forward, 5'-GTCCAC GAAGCAGCGCAAATGCCAGCAGTGTACG-3'; reverse, 5'-GCTGACACT GCTGGCATTTCGCTGCTTCGTGGAC-3'.

PCR was performed with one set of primers at a time. DpnI was added to the PCR product for 2 h at 37 °C before transformation of DH5a bacteria. Plasmid preps were sequenced before the next mutation introduced.

PSMD11-overexpressing lentiviral construct (PSMD11(OE)) was generated as follows. Human PSMD11 complementary DNA was PCR-amplified and cloned into pLVX-Puro using XhoI and BamHI. Resulting constructs were transformed into One Shot Stbl3 *Escherichia coli* (Invitrogen). Constructs were sequence verified and thereafter transfected into packaging cells to produce high titre lentiviruses.

Lentiviral infection of human stem cells. hESC colonies growing on Matrigel were incubated with mTeSR1 medium containing 10 μ M ROCK inhibitor for 1 h and individualized using Accutase. Fifty-thousand cells were infected in suspension with 10 μ l of concentrated lentivirus in the presence of 10 μ M ROCK inhibitor. Cell suspension was centrifuged to remove virus, passed through a mesh of 40 μ M to obtain individual cells, and plated back on a feeder layer of fresh MEFs in hESC media supplemented with 10 μ M ROCK inhibitor. After a few days in culture, small hESC colonies arose. For the GFP and FOXO shRNA lentiviral stable lines, GFP-positive colonies were selected and manually passaged onto fresh MEFs to establish new hESC lines. For non-targeting shRNA lentivirus, PSMD11 shRNA, PSMC2 shRNA, FOXO4(OE) shRNA, FOXO4-AAA(OE) shRNA and PSMD11(OE) shRNA stable lines, we performed 1 μ g ml⁻¹ puromycin-resistance selection during three days and then colonies were manually passaged onto fresh MEFs to establish new hESC lines.

Transient infection experiments were performed as follows. hESC colonies growing on Matrigel were incubated with mTeSR1 medium containing 10 μ M ROCK inhibitor for 1 h and individualized using Accutase. Ten thousand cells were plated on Matrigel plates and incubated with mTeSR1 medium containing 10 μ M ROCK inhibitor for 1 day. Cells were infected with 2 μ l of concentrated lentivirus. Plates were centrifuged at 800g for 1 h at 30 °C. Cells were fed with fresh media the day after to remove virus. NPCs were split as described above, and infected with 2 μ l of concentrated lentivirus for 1 day. Neurons were infected after 2 months of differentiation with 2 μ l of concentrated lentivirus for 1 day. In all the cases, cells were collected for experimental assays after four days of infection.

26S proteasome fluorogenic peptidase assays. The *in vitro* assay of 26S proteasome activities was performed as previously described¹⁰. Cells were collected in proteasome activity assay buffer (50 mM Tris-HCl, pH 7.5, 250 mM sucrose, 5 mM MgCl₂, 0.5 mM EDTA, 2 mM ATP and 1 mM dithiothreitol) and lysed by passing ten times through a 27-gauge needle attached to a 1-ml syringe. Lysate was centrifuged at 10,000g for 10 min at 4 °C. Approximately 15–25 μ g of total protein of cell lysates were transferred to a 96-well microtiter plate (BD Falcon), and the fluorogenic substrate was then added to lysates. To measure the chymotrypsin-like activity of the proteasome we used either Z-Gly-Gly-Leu-AMC (Enzo) or Suc-Leu-Leu-Val-Tyr-AMC (Enzo). We used Z-Leu-Leu-Glu-AMC (Enzo) to measure the caspase-like activity of the proteasome, and Ac-Arg-Leu-Arg-AMC for the proteasome trypsin-like activity. Fluorescence (380 nm excitation, 460 nm emission) was monitored on a microplate fluorometer (Infinite M1000, Tecan) every 5 min for 1 h at 37 °C. Protein concentration of the cell homogenates was determined using the BCA protein assay (Pierce). For SDS experiments, 0.025% SDS was added to cell lysates 5 min before the digestion assay.

Native gel immunoblotting of the proteasome. hESCs (H9), NPCs and neurons were collected in proteasome activity assay buffer (50 mM Tris-HCl, pH 7.6, 5 mM MgCl₂, 0.5 mM EDTA, 5 mM ATP, 1 mM dithiothreitol and 10% glycerol supplemented with Roche phosphatase inhibitors) and lysed by passing ten times

through a 27-gauge needle attached to a 1-ml syringe. Lysate was centrifuged at 16,000g for 15 min at 4 °C. Fifteen micrograms of total protein was run on a 3–12% NativePAGE Bis-Tris gel (Invitrogen) in NativePAGE running buffer (Invitrogen) at 4 °C for 1 h at 150 V and then increased to 200 V for a further hour. Proteins were then transferred to a polyvinylidene difluoride (PVDF) membrane at 25 V for 1 h in NativePAGE transfer buffer (Invitrogen) in an XCell II Blot module (Invitrogen). After transfer, the PVDF membrane was incubated for 20 min with 8% acetic acid to fix the proteins and then dried. Western blot analysis was performed with anti-20S alpha 1–7 (Abcam) and anti-PSMD2 (Abcam).

HEK293T cells were run on 3.5% native gels prepared in resolving buffer (90 mM Tris base, 90 mM boric acid, 5 mM MgCl₂, 0.5 mM EDTA and 1 mM ATP) with 5 mM ATP, 1 mM dithiothreitol and 3.5% acrylamide from a 40% stock solution of acrylamide and bisacrylamide in a 37.5:1 ratio (Bio-Rad). These were run at 110 V for 3 h at 4 °C. Activity assays were performed by incubating the gels in activity assay buffer for 20 min at 37 °C and developed using a BioRad Gel Doc with ultraviolet illumination. Before transfer, the gels were incubated in transfer buffer (25 mM Tris base and 192 mM glycine) with 1% SDS for 10 min followed by a 10-min incubation in transfer buffer. The protein was transferred to PVDF membranes at 5 V for 16 h in transfer buffer using an Idea Scientific Genie Blotter. Western blot analysis was performed with anti-PSMD1 (Abcam) and analysed using the Odyssey system (LI-COR Biosciences). Extracts were also analysed by SDS–PAGE to determine protein expression levels and loading control.

Western blot. For analysis of proteasome subunits, cells were collected in proteasome activity assay buffer supplemented with protease inhibitors (Roche) and lysed by passing ten times through a 27-gauge needle attached to a 1-ml syringe. Lysate was centrifuged at 10,000g for 10 min at 4 °C. Protein concentration of the cell homogenates was determined using the BCA protein assay (Pierce). For analysis of transcription factor and polyubiquitinated proteins, cells were collected from tissue culture plates by cell scraping and lysed in protein cell lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM EDTA, 50 mM NaCl, 50 mM NaF, 1% Triton X-100, 0.1% SDS supplemented with 2 mM sodium orthovanadate, 1 mM phenylmethylsulphonyl fluoride and Complete mini protease and PhosSTOP inhibitor cocktail mix) for 2 h at 1,000 r.p.m. and 4 °C in a Thermomixer. Protein concentrations were determined with a standard Bradford protein assay (BioRad). Approximately 20–50 µg of total protein was separated by SDS–PAGE, transferred to nitrocellulose membranes (Whatman) and subjected to immunoblotting. Western blot analysis was performed with anti-FK1 (Enzo), anti-20S alpha 1–7 (Abcam), anti-proteasome 20S C2 (Abcam), anti-Rpt6 (Biomol), anti-PSMD1 (Abcam), anti-PSMD2 (Abcam), anti-PSMD14 (Abcam), anti-PSMB6 (Abcam), anti-PSMD11 (Novus), anti-FOXO4 (55D4)

(Cell Signaling), anti-FOXO1 (C29H4) (Cell Signaling), anti-SOX2 (D6D9) (Cell Signaling), anti-FGF5 (Abcam), anti-MSX1 (Abcam), anti-PAX6 (Abcam) and anti-β-actin (Abcam). The affinity of the antibody to PSMD11 has been characterized by detecting a decrease at the protein levels with *PSMD11* shRNA or an increase by ectopic expression of *PSMD11*. These experiments convincingly show differences in only one band and we ascribe any alteration of *PSMD11* to this band.

Coomassie staining. Protein lysates were separated by SDS–PAGE and visualized directly in the gel by Coomassie staining²⁶. Gels were incubated in fixing solution (50% methanol, 10% acetic acid and 100 mM ammonium acetate) for 60 min, stained with 0.025% Coomassie dye in 10% acetic acid on a shaker overnight and destained twice in 10% acetic acid for 60 min. Gels were transferred to water and analysed with the Odyssey imager (Li-Cor Bioscience).

Bromodeoxyuridine proliferation assay. Cells were incubated with media containing 10 mM bromodeoxyuridine (BrdU) for 40 min. Cells were fixed with formaldehyde 4% in PBS for 15 min and washed in PBS. Before permeabilization, cells were incubated for 1 h in 2 N HCl at room temperature followed by extensive washes in PBS. Cells were permeabilized with 0.5% Triton X-100 in PBS for 10 min and blocked with 5% normal donkey serum in 1% PBS-BSA for 40 min at room temperature. Rabbit anti-BrdU antibody (ABD Serotech) was diluted in 1% PBS-BSA and used for overnight incubation followed by incubation with a biotinylated anti-rabbit secondary antibody (Vector) for a further 2 h at room temperature. Finally, cells were incubated with streptavidin-AlexaFluor 568 (Jackson Immuno Research) for 1 h. DAPI was used to visualize nuclei at a concentration of 0.5 µg ml⁻¹ in PBS.

RNA isolation and quantitative RT–PCR. Total RNA was extracted using RNAbec (Tel-Test Inc.). cDNA was created using the Quantitect reverse transcriptase kit (Qiagen). SybrGreen real-time qPCR experiments were performed as described in the manual using ABI Prism 7900HT (Applied Biosystems) and cDNA at a 1:20 dilution. Data was analysed with the comparative 2 $\Delta\Delta C_t$ method using β-actin and *GAPDH* as housekeeping genes. See Supplementary Table 9 for details about the primers used for this assay.

23. Xu, R. H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature Biotechnol.* **20**, 1261–1264 (2002).
24. Itoh, M., Kiuru, M., Cairo, M. S. & Christiano, A. M. Generation of keratinocytes from normal and recessive dystrophic epidermolysis bullosa-induced pluripotent stem cells. *Proc. Natl Acad. Sci. USA* **108**, 8797–8802 (2011).
25. Tiscornia, G., Singer, O. & Verma, I. M. Design and cloning of lentiviral vectors expressing small interfering RNAs. *Nature Protocols* **1**, 234–240 (2006).
26. Schägger, H. Tricine-SDS-PAGE. *Nature Protocols* **1**, 16–22 (2006).

Quantum teleportation over 143 kilometres using active feed-forward

Xiao-Song Ma^{1,2,†}, Thomas Herbst^{1,2}, Thomas Scheidl¹, Daqing Wang¹, Sebastian Kropatschek¹, William Naylor¹, Bernhard Wittmann^{1,2}, Alexandra Mech^{1,2}, Johannes Kofler^{1,3}, Elena Anisimova⁴, Vadim Makarov⁴, Thomas Jennewein^{1,4}, Rupert Ursin¹ & Anton Zeilinger^{1,2}

The quantum internet¹ is predicted to be the next-generation information processing platform, promising secure communication^{2,3} and an exponential speed-up in distributed computation^{2,4}. The distribution of single qubits over large distances via quantum teleportation⁵ is a key ingredient for realizing such a global platform. By using quantum teleportation, unknown quantum states can be transferred over arbitrary distances to a party whose location is unknown. Since the first experimental demonstrations of quantum teleportation of independent external qubits⁶, an internal qubit⁷ and squeezed states⁸, researchers have progressively extended the communication distance. Usually this occurs without active feed-forward of the classical Bell-state measurement result, which is an essential ingredient in future applications such as communication between quantum computers. The benchmark for a global quantum internet is quantum teleportation of independent qubits over a free-space link whose attenuation corresponds to the path between a satellite and a ground station. Here we report such an experiment, using active feed-forward in real time. The experiment uses two free-space optical links, quantum and classical, over 143 kilometres between the two Canary Islands of La Palma and Tenerife. To achieve this, we combine advanced techniques involving a frequency-uncorrelated polarization-entangled photon pair source, ultra-low-noise single-photon detectors and entanglement-assisted clock synchronization. The average teleported state fidelity is well beyond the classical limit⁹ of two-thirds. Furthermore, we confirm the quality of the quantum teleportation procedure without feed-forward by complete quantum process tomography. Our experiment verifies the maturity and applicability of such technologies in real-world scenarios, in particular for future satellite-based quantum teleportation.

Significant progress has been made recently in the field of quantum communication based on optical free-space links^{10–19}, which potentially allow much larger propagation distances compared to the existing fibre networks because of the lower photon loss per kilometre. To enable quantum communication on a global scale and among parties not having access to any fibre network, it is foreseeable that experiments will involve ground-to-satellite^{10,11,15} and inter-satellite links.

Previous experiments focused on the distribution of quantum states of single photons or entangled photon pairs over optical free-space links. However, the realization of more sophisticated multiphoton quantum information protocols, such as quantum teleportation, remained an experimental challenge under real-world conditions. Quantum teleportation is based on the simultaneous creation of at least three photons, which for random photon pair sources reduces the count rate by several orders of magnitude compared to experiments using only two photons^{16,19}. This decreases the signal-to-noise ratio and requires a long integration time, such that high system

stability is necessary. Moreover, the complexity and environmental requirements of a quantum teleportation set-up are increased significantly compared to previous two-photon experiments, which provides significant experimental and technological challenges. The work presented in ref. 18 was a significant achievement in long-distance quantum communication, harnessing the entanglement between different degrees of freedom of a single photon. However, because the teleported qubit was not provided independently from the outside, the applicability of that scheme is limited. Most earlier experiments on teleportation of qubits and squeezed states^{6–8} were in-laboratory demonstrations, and hence the communication distances were rather short. Although fibre-based teleportation has been demonstrated experimentally^{20,21}, the maximum transmission distance is limited by the intrinsic photon loss in optical fibre, unless quantum repeaters are involved²². In comparison to these previous studies, the experiment presented in this Letter achieves long-distance free-space teleportation of an independent quantum state, thus paving the way for satellite-based global quantum communication.

Quantum teleportation relies on using both a quantum channel and a classical channel between two parties, usually called Alice and Bob⁵ (here located in La Palma and Tenerife respectively; Fig. 1a). The quantum channel is used by Alice and Bob to share the entangled auxiliary state²³

$$|\Psi^-\rangle_{23} = \frac{1}{\sqrt{2}}(|H\rangle_2|V\rangle_3 - |V\rangle_2|H\rangle_3) \quad (1)$$

which is one of the four maximally entangled Bell states ($|\Psi^\pm\rangle = \frac{1}{\sqrt{2}}(|H\rangle|V\rangle \pm |V\rangle|H\rangle)$ and $|\Phi^\pm\rangle = \frac{1}{\sqrt{2}}(|H\rangle|H\rangle \pm |V\rangle|V\rangle)$). $|H\rangle$ and $|V\rangle$ denote the horizontal and vertical polarization states. Alice and Bob share this entangled state, where photon 2 is with Alice and photon 3 is with Bob. Charlie provides the input photon 1 to be teleported to Alice in a general polarization state

$$|\phi\rangle_1 = \alpha|H\rangle_1 + \beta|V\rangle_1 \quad (2)$$

where α and β are complex numbers ($|\alpha|^2 + |\beta|^2 = 1$), unknown to both Alice and Bob.

Alice then performs a Bell-state measurement (BSM), projecting photons 1 and 2 randomly onto one of the four Bell states each with the same probability of 25%. As a consequence, photon 3 is projected onto the input state $|\phi\rangle$, up to a unitary transformation (U), which depends on the outcome of the BSM. When Alice feeds the outcome of the BSM forward to Bob via the classical channel, he can implement the corresponding unitary operation in real time and thus obtain photon 3 in the initial state (equation (2)) of photon 1. If $|\Psi^-\rangle_{12}$ is detected, then U corresponds to the identity operation, which means that Bob needs to do nothing. If, on the other hand, $|\Psi^+\rangle_{12}$ is detected, Bob has to apply a π phase shift between the horizontal and the vertical component of his photon 3.

¹Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, A-1090 Vienna, Austria. ²Vienna Center for Quantum Science and Technology, Faculty of Physics, University of Vienna, Boltzmanngasse 5, A-1090 Vienna, Austria. ³Max Planck Institute of Quantum Optics, Hans-Kopfermann-Straße 1, 85748 Garching/Munich, Germany. ⁴Institute for Quantum Computing and Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada. [†]Present address: Department of Electrical Engineering, Yale University, New Haven, Connecticut 06520, USA.

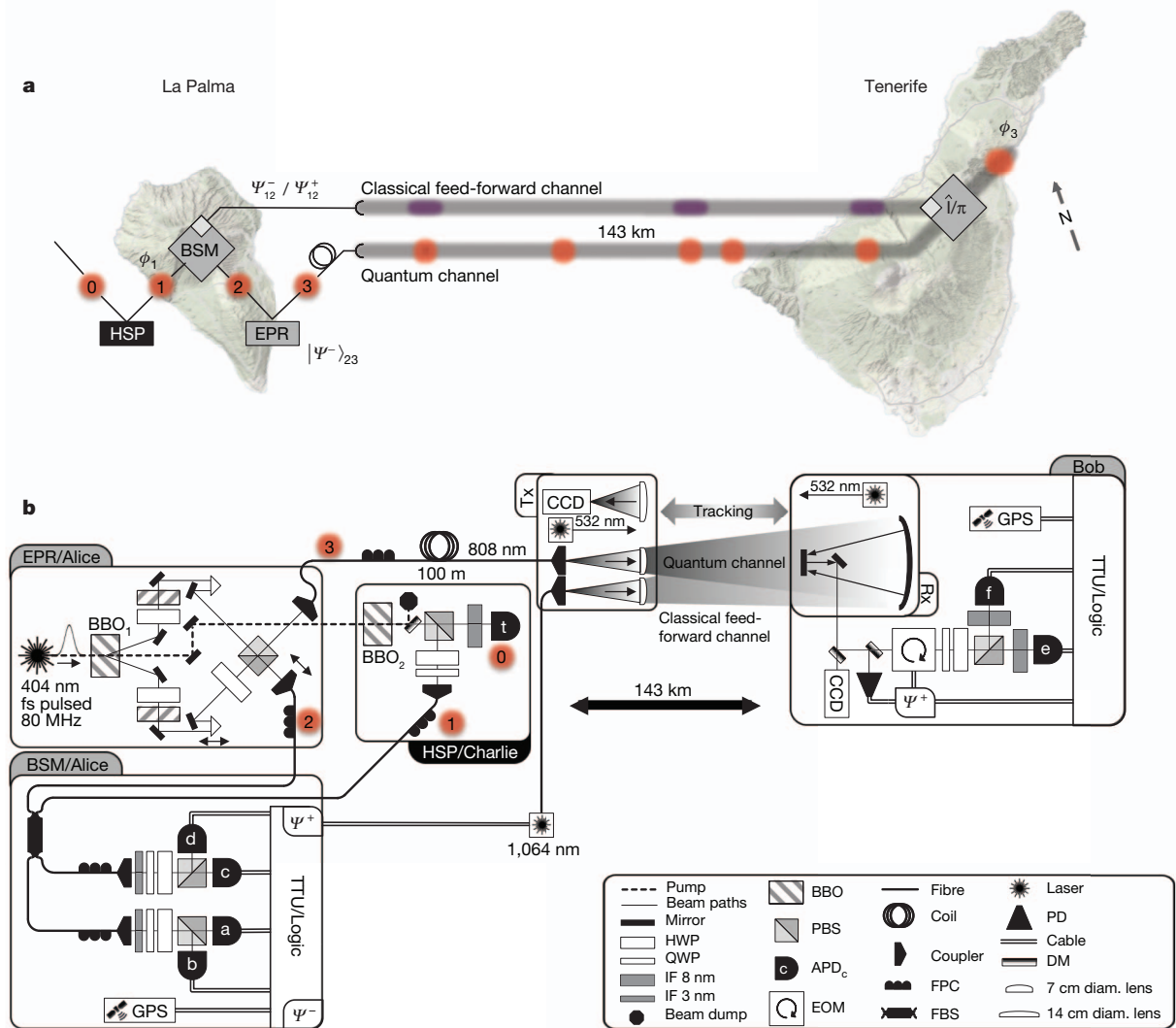


Figure 1 | Quantum teleportation between the Canary Islands La Palma and Tenerife over both quantum and classical 143-km free-space channels.

a, Experimental scheme. Alice and Charlie are situated in La Palma, and Bob in Tenerife. Charlie prepares the teleportation input photon 1 in $|\phi\rangle_1$, using a heralded single-photon (HSP) source with a trigger photon 0 (photons are indicated by black numerals on red circles). An Einstein–Podolsky–Rosen (EPR)²³ source generates an entangled pair of photons 2 and 3 in the state $|\Psi^-\rangle_{23}$. Alice then performs a Bell-state measurement (BSM) on photons 1 and 2, and projects them onto two of the four Bell states ($|\Psi^-\rangle_{12}/|\Psi^+\rangle_{12}$), and sends the result via the classical feed-forward channel to Bob. Photon 3 is sent via the free-space quantum channel to Bob, who applies a unitary transformation (identity operation or π phase shift) on photon 3 depending on the BSM result and thus turns its state $|\phi\rangle_3$ into a replica of the initial quantum state $|\phi\rangle_1$. **b**, Set-up. In La Palma, a frequency-uncorrelated polarization-entangled

Our experiment was conducted between Alice’s transmitter station (at the Jacobus Kapteyn Telescope (JKT) of the Isaac Newton Group on La Palma) and Bob’s receiver station (at the Optical Ground Station (OGS) of the European Space Agency on Tenerife), separated by 143 km, both at altitudes of about 2,400 m. Our experimental set-up is shown in Fig. 1b. At La Palma, near-infrared femtosecond pulses (with a central wavelength of 808 nm) emitted from a mode-locked Ti:sapphire laser, with a repetition rate of 80 MHz, were up-converted to blue pulses (with a central wavelength of 404 nm). They were used to generate two photon pairs via type-II spontaneous parametric down conversion (SPDC) in two nonlinear β -barium borate (BBO) crystals placed in sequence. The first SPDC source was aligned to emit the entangled auxiliary photon pairs (photons 2 and 3) in the $|\Psi^-\rangle_{23}$ state²⁴, equation (1), while the second crystal was a heralded

photon pair source generated photons 2 and 3 in BBO₁ (EPR/Alice) and a collinear photon pair source generated photons 0 and 1 in BBO₂ (HSP/Charlie). All single photons were coupled into single-mode fibres. For implementing the BSM, photons 1 and 2 interfered in a fibre beam splitter (FBS) followed by polarization-resolving single-photon detection (BSM/Alice). Photon 3 was guided to the transmitter telescope via a 100-m single-mode fibre and sent to Bob in Tenerife, where the unitary transformation was implemented using an electro-optical modulator (EOM) and photon 3’s polarization was measured. A real-time feed-forward operation was implemented by encoding the $|\Psi^+\rangle_{12}$ BSM result in 1,064-nm laser pulses, which were then sent to Bob via the feed-forward channel. On Bob’s side, they were separated by a dichroic mirror (DM), detected with a photodetector (PD) and used to trigger the EOM to perform the required π phase shift operation. See main text for details.

single-photon (HSP) source providing Charlie’s photon 1 to be teleported. That source delivered pairs of horizontally (photon 0) and vertically (photon 1) polarized photons in a product state. The detection of photon 0 by the avalanche photodiode (APD) *t* served as a trigger to herald the presence of photon 1. All photons were spectrally filtered using interference filters (IF) and coupled into single-mode fibres for spectral and spatial mode selection.

To realize the BSM, Alice’s photon 2 was overlapped on a fibre beam splitter (FBS) with the teleportation input photon 1, whose polarization was arbitrarily prepared by Charlie using half- and quarter-wave plates (HWP and QWP). In each output port of the FBS, a polarizing beam splitter (PBS) was used to project these photons on either horizontal or vertical polarization. Fibre polarization controllers (FPCs) were used to compensate for unwanted polarization rotation induced

by the fibres. Our BSM can identify two out of the four Bell states, which is the optimum achievable with linear optics only²⁵. For more details on our BSM, see Supplementary Information. While the BSM was being performed at Alice, photon 3 was guided to a 7-cm-aperture transmitter telescope through a 100-m-long single-mode fibre and then sent via a 143-km free-space quantum channel over to Bob in Tenerife. There, it was collected by the 1-m-aperture OGS telescope, and guided through its Coudé path to Bob.

In the first stage of our experiment, we only considered the cases where Alice detected $|\Psi^-\rangle_{12}$ in the BSM, which results in photon 3 being already in the state of the input photon, $|\phi\rangle_1$, and hence Bob was required to perform an identity operation, that is, to do nothing at all. We verified the success of the teleportation process by analysing the polarization state of photon 3, which was accomplished by a polarization analyser, consisting of a quarter- and a half-wave plate and two free-space coupled Si-APDs (silicon avalanche photodiodes) placed in each output mode of a polarizing beam splitter.

In the second stage of our experiment, we implemented a real-time feed-forward operation. When Alice obtained $|\Psi^+\rangle_{12}$, she sent this classical information to Bob. On receiving this information, Bob had to apply a π phase shift between the $|H\rangle$ and $|V\rangle$ components of photon 3 to obtain the replica of the input state $|\phi\rangle_1$. For further details on the feed-forward implementation, see Supplementary Information.

The relevant events on Alice's and Bob's sides were recorded with separate time-tagging units each disciplined to the global positioning system (GPS)¹⁴. First, Alice in La Palma identified the threefold coincidence events corresponding to the $|\Psi^\pm\rangle_{12}$ outcomes of the BSM. This was done using a coincidence logic circuit featuring two separate output signals realized with transistor-transistor logic (TTL) pulses. These TTL pulses were fed into a time-tagging unit which recorded the exact time and the BSM result into a binary file. Similarly, Bob fed either the signals of both detectors (stage 1, without feed-forward) or the coincidence between these signals and the BSM results sent via 1,064-nm laser pulses (stage 2, with feed-forward) into his time-tagging unit. After a measurement run was completed, both time-tagged data files were compared by cross-correlation, and the detection events associated with simultaneous detection of four photons originating from the same pump pulse were identified.

The real-life long-distance environment provided a number of challenges for the present teleportation experiment. These challenges resulted most significantly in the need to cope with an extremely low signal-to-noise ratio when using standard techniques—indeed, too low to perform a successful experiment. To enhance the signal-to-noise ratio to a level that made the experiment possible, we used the following advanced techniques: a frequency-uncorrelated polarization-entangled photon pair source^{26–28}, ultra-low-noise large-active-area single-photon detectors at Bob²⁹, and entanglement-assisted clock synchronization^{13,14,17}. Whereas all these techniques have been implemented individually, our work is (to our knowledge) the first that combines all of them simultaneously and moreover in an outdoor environment. See Supplementary Information for details.

First we present our results without feed-forward, where we only considered the BSM outcome $|\Psi^-\rangle_{12}$. The input state $|\phi\rangle_1$ was always approximately one of the four ideal input states $|\phi_{\text{ideal}}\rangle \in \{|H\rangle, |V\rangle, |P\rangle = (|H\rangle + |V\rangle)/\sqrt{2}, |L\rangle = (|H\rangle - i|V\rangle)/\sqrt{2}\}$. We performed tomographic measurements on three consecutive nights, thereby accumulating data over 6.5 h. Figure 2 shows the state tomography results of quantum teleportation. The measured density matrix ρ for each of these teleported states was reconstructed from the experimentally obtained data using the maximum-likelihood technique³⁰. The fidelity of the teleported state is defined as the overlap of the ideal teleported state $|\phi_{\text{ideal}}\rangle$ with the measured density matrix: $f = \langle\phi_{\text{ideal}}|\rho|\phi_{\text{ideal}}\rangle$. For this set of states, the teleported state fidelities are measured to be $f = 0.890(42)$, $0.865(46)$, $0.845(27)$ and $0.852(37)$, yielding an average $f = 0.863(38)$, where digits in parentheses represent 1σ uncertainties, for example, $f = 0.863 \pm 0.038$. During these measurements the link attenuation varied from 28.1 dB to 39.0 dB, which was mainly caused by rapid temperature change and strong wind. Despite such high loss in the quantum free-space channel, the classical average fidelity limit⁹ of $2/3$ was clearly surpassed by our observed fidelities, as shown in Fig. 3. (Note that our random sampling of the input states over the mutually unbiased bases states leads to the same classical limit as sampling over the whole Bloch sphere. Also, it does not matter whether the sampling is done over all six mutually unbiased basis states or over three states with one state per basis.) Therefore, we have explicitly demonstrated quantum teleportation over the 143-km free-space channel.

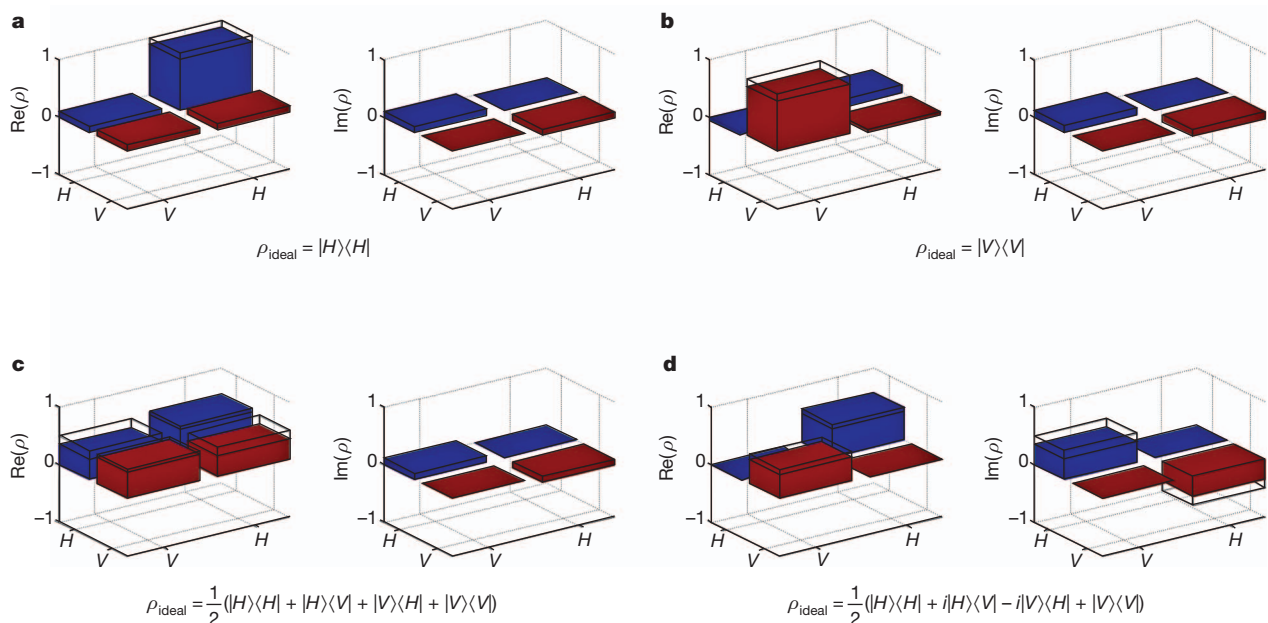


Figure 2 | State tomography results of the four quantum states without feed-forward over the 143-km free-space channel with the BSM outcome of $|\Psi^-\rangle_{12}$. The bar graphs show the reconstructed density matrices ρ for the four states teleported from Alice (La Palma) to Bob (Tenerife) over the 143-km free-

space channel. The wire grid indicates the expected values for the ideal cases. The data shown comprise a total of 605 fourfold coincidence counts in about 6.5 h. The uncertainties in state fidelities extracted from these density matrices are calculated using a Monte Carlo routine assuming Poissonian errors.

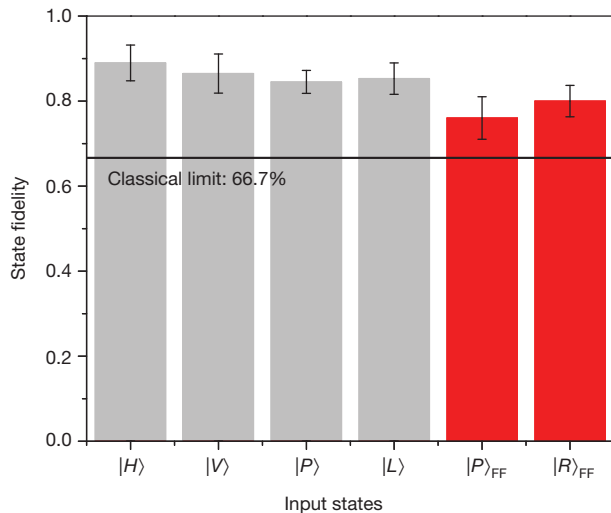


Figure 3 | Summary of the state fidelity results for the teleported quantum states with and without feed-forward. Grey bars show the results obtained in quantum teleportation without feed-forward. Bob was informed via an internet connection when Alice's BSM outcome was $|\Psi^-\rangle_{12}$, that is, those cases where he is not required to apply any operation to his photon. The uncertainties of these state fidelities are the same as presented in the main text. Red bars stand for the results obtained in quantum teleportation with feed-forward (FF). Bob was informed via the classical free-space link of Alice's BSM outcomes of $|\Psi^-\rangle_{12}$. The uncertainties of these state fidelities are derived from Poissonian statistics. All observed fidelities significantly exceed the classical average fidelity limit of $2/3$ (66.7%). Error bars, $\pm 1\sigma$.

The reconstructed density matrices of the teleported quantum states allow us to fully characterize the teleportation procedure by quantum process tomography. The four input states ($\rho_{\text{ideal}} = |\phi_{\text{ideal}}\rangle\langle\phi_{\text{ideal}}| = |H\rangle\langle H|, |V\rangle\langle V|, |P\rangle\langle P|, |L\rangle\langle L|$) are transferred to the corresponding (reconstructed) output states ρ . We can completely describe the effect of teleportation on the input states ρ_{ideal} by determining the process matrix χ , defined by $\rho = \sum_{l,k=0}^3 \chi_{lk} \sigma_l \rho_{\text{ideal}} \sigma_k$, where the σ_i are the Pauli matrices with σ_0 the identity operator. The process matrix χ can be computed analytically from these four equations for the four different input and output states². The ideal process matrix of quantum teleportation χ_{ideal} has only one non-zero component, $(\chi_{\text{ideal}})_{00} = 1$, meaning the input state is teleported without any reduction in quality. Figure 4a and b show respectively the real and imaginary components of χ for quantum teleportation based on our experimental results. The process fidelity of our experiment was $f_{\text{process}} = \text{Tr}(\chi_{\text{ideal}}\chi) = 0.710(42)$. This clearly confirmed the quantum nature of our teleportation experiment as it is five standard deviations above the maximum process

fidelity of 0.5, which is the limit one can reach with a classical strategy where Alice and Bob do not share any entanglement as a resource.

In the second stage of the experiment, we realized quantum teleportation including real-time feed-forward of the BSM result over the 143-km classical channel. We set $|P\rangle$ and $|R\rangle$ states ($|R\rangle = (|H\rangle + i|V\rangle)/\sqrt{2}$) as input states for which the required π phase shift between the $|H\rangle$ and the $|V\rangle$ components of photon 3 resulted in a 90° polarization rotation. However, for the $|H\rangle$ or $|V\rangle$ input state, feed-forward is irrelevant because a π phase shift would only result in a non-detectable global phase shift. Thus, the quality of teleportation of these states is already confirmed by our first-stage experiment. Realizing teleportation for the states $|P\rangle$ and $|R\rangle$ fully confirms the generality of the procedure, as these states belong to different mutually unbiased bases. In Tenerife, we analysed photon 3 in the eigenbasis of the input state, that is, the $|P\rangle/|M\rangle$ ($|R\rangle/|L\rangle$) basis when the input state was $|P\rangle$ ($|R\rangle$). Here $|M\rangle = (|H\rangle - |V\rangle)/\sqrt{2}$. The resultant fidelities of the teleported states are 0.760(50) and 0.800(37) for $|P\rangle$ and $|R\rangle$, respectively (red bars in Fig. 3). Both results are clearly above the classical fidelity bound. Note that in our experiment, the efficiency of the classical link was measured to be 21.3%, which was mainly due to amplitude fluctuations caused by atmospheric turbulence.

Using the real-time feed-forward operation, we unambiguously experimentally demonstrated quantum teleportation from La Palma to Tenerife over a 143-km free-space channel. We note that from a conceptual perspective, real-time feed-forward is part of the original teleportation proposal⁵. Ultimately, the advantage of long-distance teleportation compared to just sending the quantum state itself may lie in the following future applications: if Alice and Bob can stockpile their entangled states beforehand (with the help of quantum memories), teleportation is advantageous if the quantum channel is of low quality or if Bob's location is unknown to Alice. This is because Alice can broadcast the classical information with high quality and to arbitrary, or even unknown, locations⁵. (Also, the quantum repeater, which is of high importance for large-scale quantum networks, is based on teleportation in the form of entanglement swapping³¹.) Specifically, this advantage also shows the usefulness of quantum teleportation even for quantum channels that allow only small transmission rates. We note that over the years teleportation sources have been markedly improved. We believe this development will continue and currently several new schemes are being pursued.

Our work proves the feasibility of both ground-based and satellite-based free-space quantum teleportation. Our quantum teleportation set-up was able to achieve coincidence production rates and fidelities sufficient to cope with the optical link attenuation resulting from various experimental and technical challenges; such challenges will arise in quantum transmission between a ground-based transmitter

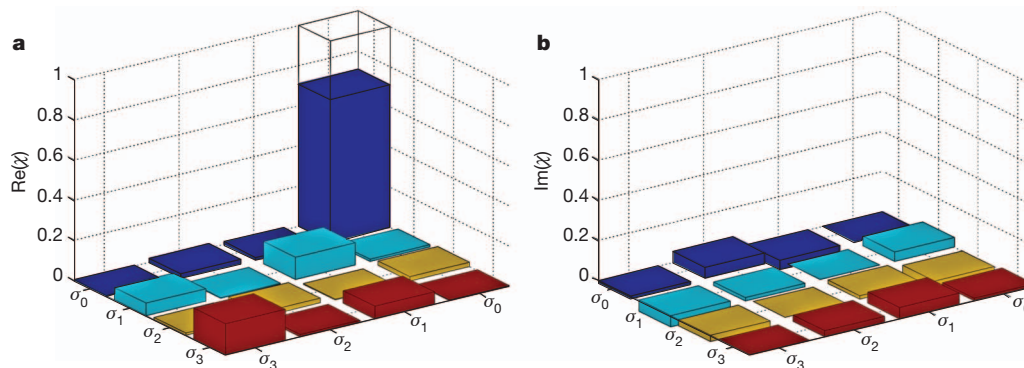


Figure 4 | Quantum process tomography of quantum teleportation without feed-forward. a, b, The real ($\text{Re}(\chi_{lk})$; a) and imaginary ($\text{Im}(\chi_{lk})$; b) values of the components of the reconstructed quantum process matrix, with $l, k = 0, 1, 2$, and 3. The results of the state tomography of the four teleported states, $|H\rangle, |V\rangle, |P\rangle, |L\rangle$, are employed to reconstruct the process matrix of quantum state

teleportation. The operators σ_i are the identity ($i = 0$) and the x -, y -, and z -Pauli matrices ($i = 1, 2, 3$). For the ideal case, the only non-zero component of the process matrix of quantum teleportation, χ_{ideal} is $(\chi_{\text{ideal}})_{00} = 1$, which is indicated by the wire grid. Our experiment clearly confirmed that χ_{00} (identity operation) is indeed the dominant component and surpasses the classical limit.

and a low-Earth-orbiting satellite receiver³². In fact, some of the demands of our experiment were even more challenging than in satellite communication since there the atmospheric distances to be overcome are certainly shorter than the distance between Tenerife and La Palma. In addition, acquiring, pointing and tracking (APT) for diffraction-limited transmitter and receiver telescopes in space is a well-established technology. Therefore, our experiment represents a crucial step towards future quantum networks in space, which require space-to-ground quantum communication. During the finalization of our manuscript, a related work was reported³³ that used a 150-Hz APT system, while our APT system operated at a few hertz, yet fast enough to fulfil the specific requirements of our experiment. The most significant distinction between ref. 33 and the experiment presented in this Letter is our implementation of an active feed-forward technique. The technology implemented in both experiments has certainly reached the required maturity for both satellite and long-distance ground communication. We expect that many of the features implemented here will be important building blocks for a new area of experiments.

Received 21 May; accepted 14 August 2012.

Published online 5 September 2012.

- Kimble, H. J. The quantum internet. *Nature* **453**, 1023–1030 (2008).
- Nielsen, M. & Chuang, I. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- Gisin, N. & Thew, R. Quantum communication. *Nature Photon.* **1**, 165–171 (2007).
- Ladd, T. D. *et al.* Quantum computers. *Nature* **464**, 45–53 (2010).
- Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
- Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
- Massar, S. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).
- Hughes, R. J. *et al.* Free-space quantum key distribution in daylight. *J. Mod. Opt.* **47**, 549–562 (2000).
- Rarity, J. G., Tapster, P. R., Gorman, P. M. & Knight, P. Ground to satellite secure key exchange using quantum cryptography. *N. J. Phys.* **4**, 82 (2002).
- Aspelmeyer, M. *et al.* Long-distance free-space distribution of quantum entanglement. *Science* **301**, 621–623 (2003).
- Marcikic, I., Lamas-Linares, A. & Kurtsiefer, C. Free-space quantum key distribution with entangled photons. *Appl. Phys. Lett.* **89**, 101122 (2006).
- Ursin, R. *et al.* Entanglement-based quantum communication over 144 km. *Nature Phys.* **3**, 481–486 (2007).
- Villoresi, P. *et al.* Experimental verification of the feasibility of a quantum channel between space and Earth. *N. J. Phys.* **10**, 033038 (2008).
- Fedrizzi, A. *et al.* High-fidelity transmission of entanglement over a high-loss free-space channel. *Nature Phys.* **5**, 389–392 (2009).
- Scheidl, T. *et al.* Feasibility of 300 km quantum key distribution with entangled states. *N. J. Phys.* **11**, 085002 (2009).
- Jin, X.-M. *et al.* Experimental free-space quantum teleportation. *Nature Photon.* **4**, 376–381 (2010).
- Scheidl, T. *et al.* Violation of local realism with freedom of choice. *Proc. Natl Acad. Sci. USA* **107**, 19708–19713 (2010).
- Marcikic, I., de Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. Long-distance teleportation of qubits at telecommunication wavelengths. *Nature* **421**, 509–513 (2003).
- Ursin, R. *et al.* Quantum teleportation across the Danube. *Nature* **430**, 849 (2004).
- Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
- Einstein, A., Podolsky, B. & Rosen, N. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47**, 777–780 (1935).
- Kwiat, P. G. *et al.* New high-intensity source of polarization-entangled photon pairs. *Phys. Rev. Lett.* **75**, 4337–4341 (1995).
- Calsamiglia, J. & Lütkenhaus, N. Maximum efficiency of a linear-optical Bell-state analyzer. *Appl. Phys. B* **72**, 67–71 (2001).
- Kim, Y.-H., Kulik, S. P., Chekhova, M. V., Grice, W. P. & Shih, Y. Experimental entanglement concentration and universal Bell-state synthesizer. *Phys. Rev. A* **67**, 010301(R) (2003).
- Poh, H. S., Lim, J., Marcikic, I., Lamas-Linares, A. & Kurtsiefer, C. Eliminating spectral distinguishability in ultrafast spontaneous parametric down-conversion. *Phys. Rev. A* **80**, 043815 (2009).
- Yao, X.-C. *et al.* Observation of eight-photon entanglement. *Nature Photon.* **6**, 225–228 (2012).
- Kim, Y.-S., Jeong, Y.-C., Sauge, S., Makarov, V. & Kim, Y.-H. Ultra-low noise single-photon detector based on Si avalanche photodiode. *Rev. Sci. Instrum.* **82**, 093110 (2011).
- White, A. G., James, D. F. V., Eberhard, P. H. & Kwiat, P. G. Nonmaximally entangled states: production, characterization, and utilization. *Phys. Rev. Lett.* **83**, 3103–3107 (1999).
- Zukowski, M., Zeilinger, A., Horne, M. A. & Ekert, A. K. “Event-ready-detectors” Bell experiment via entanglement swapping. *Phys. Rev. Lett.* **71**, 4287–4290 (1993).
- Aspelmeyer, M., Jennewein, T., Pfennigbauer, M., Leeb, W. R. & Zeilinger, A. Long-distance quantum communication with entangled photons using satellites. *IEEE J. Sel. Top. Quantum Electron.* **9**, 1541–1551 (2003).
- Yin, J. *et al.* Quantum teleportation and entanglement distribution over 100-kilometre free-space channels. *Nature* **488**, 185–188 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff of IAC: F. Sanchez-Martinez, A. Alonso, C. Warden, M. Serra and J. Carlos; and the staff of ING: M. Balcells, C. Benn, J. Rey, O. Vaduvescu, A. Chopping, D. González, S. Rodríguez, M. Abreu, L. González; J. Kuusela, E. Wille and Z. Sodnik; and J. Perdignes of the OGS and ESA. X.-S.M., T.J., R.U. and A.Z. thank S. Ramelow for discussions, P. Kolenderski for discussions on the SPDC source with the Bell-state synthesizer, S. Zotter for help during the early stages of the experiment, and R. Steinacker for meteorological advice. J.K. was supported by the EU project MALICIA. E.A. and V.M. thank C. Kurtsiefer and Y.-S. Kim for detector electronics design, J. Skaar for support, and the Research Council of Norway (grant No. 180439/V30) and Industry Canada for support. This work was made possible by grants from the European Space Agency (contract 4000104180/11/NL/AF), the Austrian Science Foundation (FWF) under projects SFB F4008 and CoQuS, and the FFG for the QTS project (no. 828316) within the ASAP 7 program. We also acknowledge support by the European Commission, grant Q-ESSENCE (no. 248095) and the John Templeton Foundation.

Author Contributions X.-S.M. conceived the research, designed and carried out the experiment, and analysed data. T.H., T.S. and D.W. carried out the experiment and analysed data. S.K., W.N., B.W. and A.M. provided experimental assistance during the early stage of the experiment. J.K. provided the theoretical analysis and analysed data. E.A. and V.M. developed the ultra-low-noise detectors. T.J. provided experimental and conceptual assistance, and conceived and developed the coincidence analysis code. R.U. conceived the research, planned and carried out the experiment and analysed data. A.Z. defined the scientific goals, conceived the research, designed the experiment and supervised the project. X.-S.M., T.H., T.S., J.K., R.U. and A.Z. wrote the manuscript with assistance from all other co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.-S.M. (Xiaosong.Ma@Univie.ac.at) or A.Z. (Anton.Zeilinger@Univie.ac.at).

of an appropriate promoter (Fig. 1d).

Mascre and colleagues' work sheds light on the heterogeneity of the proliferative cell populations in mouse epidermis, and has increased our understanding of stem-cell biology. But does this knowledge apply to human tissues? Caution is required when inferring aspects of human tissue physiology from animal data. For instance, clonal analyses of several human squamous epithelia (tissues such as the epidermis, cornea and conjunctiva) have unambiguously shown the existence of self-renewing cells endowed with stem-cell properties, as well as non-self-renewing cells with differing capacities for multiplication, including canonical transient amplifying cells^{8,9}. Both cell types participate in the regeneration of these epithelia in the clinic¹⁰. ■

Laura De Rosa and Michele De Luca are at the Center for Regenerative Medicine 'Stefano Ferrari', University of Modena and Reggio Emilia, Modena 41125, Italy.
e-mail: michele.deluca@unimore.it

1. Mascre, G. *et al.* *Nature* **489**, 257–262 (2012).
2. Potten, C. S. *Cell Tissue Kinet.* **7**, 77–88 (1974).
3. Clayton, E. *et al.* *Nature* **446**, 185–189 (2007).
4. Doupe, D. P. *et al.* *Dev. Cell* **18**, 317–323 (2010).
5. Wilson, A. *et al.* *Cell* **135**, 1118–1129 (2008).
6. Barker, N. *et al.* *Nature* **449**, 1003–1007 (2007).
7. Jaks, V. *et al.* *Nature Genet.* **40**, 1291–1299 (2008).
8. Barrandon, Y. & Green, H. *Proc. Natl Acad. Sci. USA* **84**, 2302–2306 (1987).
9. Pellegrini, G. *et al.* *J. Cell Biol.* **145**, 769–782 (1999).
10. Rama, P. *et al.* *N. Engl. J. Med.* **363**, 147–155 (2010).

ENVIRONMENTAL SCIENCE

The rainforest's water pump

An investigation of naturally occurring water recycling in rainforests finally marries the results of global climate models with observations. Alarming, it also suggests that deforestation can greatly reduce tropical rainfall. [SEE LETTER P.282](#)

LUIZ E. O. C. ARAGÃO

The humid tropics contain more than 35% of global forests, covering an area of 11,564,000 square kilometres (ref. 1). Tropical trees can extract deep soil water and pump it back to the atmosphere through a process called evapotranspiration. As a result, 25–56% of the rainfall in these regions can be recycled within the ecosystem². Large-scale climate models indicate that this mechanism — which maintains atmospheric moisture and so feeds far-inland regions with rainfall — may be weakened by the removal of existing tropical forests^{3,4}. Yet, paradoxically, an increase in local rainfall over deforested areas has been observed⁵.

On page 282 of this issue, Spracklen *et al.*⁶ report an analysis of tropical rainfall that combines models of atmospheric transport with satellite observations of rainfall and vegetation cover*. The authors conclude that, for more than 60% of pan-tropical land (that is, the tropics across all continents), air masses that have travelled over extensive vegetated surfaces can generate at least twice as much rainfall as air masses that have flowed over deforested lands. On the basis of their findings, the authors predict a potentially widespread reduction of rainfall in the Amazon basin, if deforestation

in the region were to continue at the rate⁷ that occurred from 1997 to 2002.

Sustaining high rates of global human-population growth requires an increase in farmable land, energy and timber supplies. Without adequate planning, these needs tend to be met by cutting down rainforests. Because ecological processes and human activities (such as agriculture) depend on water provision, scientists have, for at least two decades, investigated the impact of tropical deforestation on the hydrological cycle, especially in Amazonia.

Pioneering studies^{3,4} using global climate models (GCMs) found that vegetation removal can disrupt the water cycle. For example, models show that changes in land surface characteristics cause a reduction in evapotranspiration rates and so induce basin-wide decreases in Amazonian rainfall³. Another simulation shows that removing around 40% of the original forest cover could drive Amazonia into an irreversible, drier climate mode⁸.

These findings, however, are at odds with observations. Studies indicate that rainfall over deforested areas is, in fact, higher than in adjacent forests⁵. This is because increased surface heating in deforested areas induces upward air motion, reducing air pressure and drawing moist air from neighbouring forests into the openings. The moist air rises, and so generates

*This article and the paper under discussion⁶ were published online on 5 September 2012.

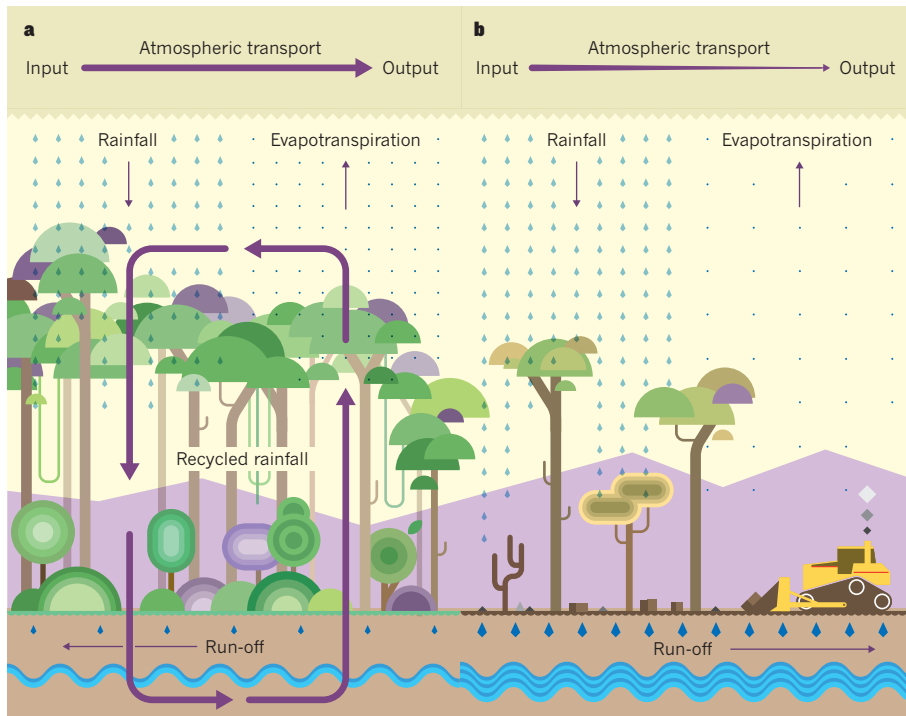


Figure 1 | Effect of deforestation on rainfall in the tropics. **a**, Much of the rainfall over tropical forests comes from water vapour that is carried by the atmosphere from elsewhere. But a large component is 'recycled' rain — water that is pumped by trees from soil into the atmosphere through a process called evapotranspiration². Water exits from forests either as run-off into streams and rivers, or as evapotranspired vapour that is carried away by the atmosphere. The atmospheric transport of water vapour into the forest is balanced by the exit of water in the form of vapour and run-off. **b**, Spracklen and colleagues' analysis⁶ suggests that deforestation reduces evapotranspiration and so inhibits water recycling. This decreases the amount of moisture carried away by the atmosphere, reducing rainfall in regions to which the moisture is transported. Decreasing evapotranspiration may also increase localized run-off and raise river levels.

convective rainfall over the openings.

The problem for scientists is that models and observations capture processes at different scales, and so are not directly comparable. But with deforestation rates reaching 50,000 square kilometres per year (an area approximately the size of England) in the humid tropics¹, resolving this issue is crucial for strategically planning the conservation of tropical biomes and for maintaining the well-being of human populations.

Enter Spracklen *et al.*, who have beautifully reconciled the scale of observations with that of models. Specifically, for each $1^\circ \times 1^\circ$ grid-cell across the pan-tropics, the authors have quantified the amount of rainfall generated by air masses that travelled over areas of varying leaf area index (LAI, the amount of leaf area per square metre of ground). The authors combined satellite observations of LAI and rainfall with the outputs of an atmospheric-transport model, which they used to estimate the trajectories of air masses in the 10 days before they arrived at each grid location. This analysis highlighted a clear connection between rainfall and the cumulative amount of LAI over which air masses travelled: for air masses exposed to low-to-medium vegetation cover, there is a nonlinear increase of 0.3 to 0.4 millimetres per

day of rainfall for every additional LAI unit, with a subsequent saturation of rainfall over densely vegetated surfaces.

Contrary to previous observations, Spracklen and colleagues' results corroborate the large-scale rainfall-driving mechanisms described by GCMs. Simply put, densely vegetated surfaces recycle water efficiently through evapotranspiration (which averages around 3–4 millimetres per day in Amazonian forests⁹), maintaining the specific humidity of the air masses over these regions (Fig. 1a). Air travelling over sparsely vegetated surfaces, however, loses moisture during continental transport because of reduced water recycling (Fig. 1b). Using a scenario of Amazonian deforestation⁷ for 2050 that assumes that the environmental policies of the early 2000s are continued into the future, the authors go on to demonstrate that forest loss is likely to reduce basin-wide rainfall by 12% during wet seasons, and by 21% during dry seasons.

The projected decline in rainfall caused by deforestation in eastern and southern Amazonia (see Fig. 4 of the paper⁶) overlaps regions that, according to GCMs, already have a high probability of increased drought frequency by the end of this century, assuming that global temperature rises by about 3°C

(ref. 10). The potential ecological and economic impacts of this overlap could be huge. Changes in regional climate could exacerbate drought-related tree mortality, which in turn would reduce carbon stocks, increase fire risk and lower biodiversity¹¹. Such changes might also directly threaten agriculture, which generates US\$15 billion per year in Amazonia¹², and the hydropower industry, which supplies 65% of Brazil's electricity¹³. Society should therefore take urgent action now, to curb tropical deforestation and avert future environmental problems.

That said, Brazil is committed by its National Plan on Climate Change to limit historical deforestation rates by 80% by the year 2020 (ref. 14), and so the scenario used by Spracklen *et al.* for their prediction is likely to overestimate the extent and magnitude of rainfall reduction in the mid-twenty-first century. An explicit consideration of temporal changes in atmospheric circulation and vegetation characteristics is also needed to refine projections of rainfall. And to provide an independent test of the patterns proposed by Spracklen *et al.*, current pan-tropical rainfall trends should be quantified.

Nevertheless, the authors have presented the first observational assessment of the influence of vegetation on tropical rainfall patterns at a spatial resolution and coverage compatible with the outputs of GCMs. Their cutting-edge methodology will allow observations to be used consistently to examine large-scale deforestation impacts on rainfall, and to refine and evaluate current models to support conservation planning in the tropics. ■

Luiz E. O. C. Aragão is at the College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4RJ, UK.
e-mail: l.aragao@exeter.ac.uk

- Hansen, M. C., Stehman, S. V. & Potapov, P. V. *Proc. Natl Acad. Sci. USA* **107**, 8650–8655 (2010).
- Eltahir, E. A. & Bras, R. L. *Rev. Geophys.* **34**, 367–378 (1996).
- Lean, J. & Warrilow, D. A. *Nature* **342**, 411–413 (1989).
- Shukla, J., Nobre, C. & Sellers, P. *Science* **247**, 1322–1325 (1990).
- Negri, A. J., Adler, R. F., Xu, L. & Surratt, J. J. *Clim.* **17**, 1306–1319 (2004).
- Spracklen, D. V., Arnold, S. R. & Taylor, C. M. *Nature* **489**, 282–285 (2012).
- Soares-Filho, B. S. *et al. Nature* **440**, 520–523 (2006).
- Sampaio, G. *et al. Geophys. Res. Lett.* **34**, L17709 (2007).
- Aragão, L. E. O. C. *et al. Geophys. Res. Lett.* **34**, L07701 (2007).
- Malhi, Y. *et al. Science* **319**, 169–172 (2008).
- Davidson, E. A. *et al. Nature* **481**, 321–328 (2012).
- Brazilian Institute for Geography and Statistics. System for Automatic Data Recovery www.sidra.ibge.gov.br (accessed August 2012).
- Brazilian Agency for Electric Energy. Energy Generation Database www.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.asp (accessed August 2012).
- Nepstad, D. *et al. Science* **326**, 1350–1351 (2009).

Diversity, stability and resilience of the human gut microbiota

Catherine A. Lozupone¹, Jesse I. Stombaugh¹, Jeffrey I. Gordon², Janet K. Jansson^{3,4} & Rob Knight^{1,5,6}

Trillions of microbes inhabit the human intestine, forming a complex ecological community that influences normal physiology and susceptibility to disease through its collective metabolic activities and host interactions. Understanding the factors that underlie changes in the composition and function of the gut microbiota will aid in the design of therapies that target it. This goal is formidable. The gut microbiota is immensely diverse, varies between individuals and can fluctuate over time — especially during disease and early development. Viewing the microbiota from an ecological perspective could provide insight into how to promote health by targeting this microbial community in clinical treatments.

Most gut microbes are either harmless or of benefit to the host. The gut microbiota protects against enteropathogens^{1,2}, extracts nutrients and energy from our diets^{3,4}, and contributes to normal immune function⁵. Disruptions to the normal balance between the gut microbiota and the host have been associated with obesity^{6,7}, malnutrition⁸, inflammatory bowel disease (IBD)^{9,10}, neurological disorders¹¹ and cancer¹². Understanding how the gut microbiota affects health and disease requires a shift in focus from individual pathogens to an ecological approach that considers the community as a whole.

The first step in understanding the symbiotic relationship between gut microbes and their host is to characterize the baseline healthy microbiota and the differences that are associated with disease. Large-scale projects such as the European Metagenomics of the Human Intestinal Tract (MetaHIT)¹³ and the US Human Microbiome Project (HMP)¹⁴ have made substantial progress towards this goal. Once the healthy composition and functional states of gut microbiota are understood, the features that, when disrupted, are associated with disease can be determined. However, the complexity of the microbiota, and the variation between and within individuals complicates the definition of what this ideal state may be within a population or an individual.

Ecological principles can aid in understanding the host–microbe interactions and specific functions of the gut microbiota. Improvements in sequencing methods and technologies such as proteomics and metabolomics, coupled with metabolic network modelling^{15,16}, show how host and environmental factors can affect gut microbial ecology over a lifetime. The composition, diversity and function of gut microbial communities have the potential to inform personalized nutritional and drug treatment strategies (Fig. 1). In this Review, we summarize the progress that has been made towards characterizing the diversity and function of microbial communities in the healthy human gut, describe the ways in which this ecosystem can go awry and discuss the prospects for restoring a degraded ecosystem.

Taxonomic diversity in the healthy gut

Until the advent of higher-throughput sequencing, even basic questions — such as how many different species there are, how diverse their genetic make-up is, which features are ubiquitous and which

are unique to an individual, and whether community function can be predicted from species present — remained unanswered.

Culture-based studies suggest that all healthy adults share most of the same gut bacterial species — a core microbiota. For example, *Escherichia coli* can be isolated from most people. However, culture-independent sequencing studies (Fig. 2) have demonstrated a vast microbial diversity that is highly variable over time and across populations. Each of us harbours more than 1,000 ‘species-level’ phylotypes: clusters of sequences that have as much diversity in their small subunit ribosomal RNA genes as named species¹⁷. Most of these phylotypes are bacteria belonging to just a few phyla. In adults, Bacteroidetes and Firmicutes usually dominate the microbiota, whereas Actinobacteria, Proteobacteria and Verrucomicrobia, although found in many people, are generally minor constituents¹⁸ (Fig. 3). Methanogenic archaea (mainly *Methanobrevibacter smithii*), eukaryotes (mainly yeasts) and viruses (mainly phage) are also present¹⁹. Despite the consistency of these main components, their relative proportions and the species present vary markedly across individuals (Figs 3 and 4). Attempts to find a core set of species-level phylotypes in the adult microbiota have identified several main players, including *Faecalibacterium prausnitzii*, *Roseburia intestinalis* and *Bacteroides uniformis*¹³, but in some individuals even these can represent less than 0.5% of the microbes present²⁰. As the data set expands to include developing countries and a broader age range^{4,21}, the concept that there is a core set of species in the microbiota is becoming more unlikely.

Functional diversity in the healthy gut

Knowing the composition of the microbial community alone does not necessarily lead to an understanding of its function. Functional information comes, in part, from studying cultured isolates that have well-characterized genome content and *ex vivo* phenotypes, and from sequencing community DNA. Functional screening by shotgun metagenomics relies on the sequencing of the total microbial community DNA, including from uncultured members, and matching the sequences to known functional genes (Fig. 2). Identifying the genes involved in specific metabolic pathways can predict functional capabilities, but without messenger RNA, protein and metabolite profiling these are still predictions. The process of mapping short gene

¹Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado, USA. ²Center for Genome Sciences and Systems Biology, Washington University in St. Louis, St. Louis, Missouri, USA; ³Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; ⁴Joint Genome Institute, Lawrence Berkeley National Laboratory, Walnut Creek, California, USA; ⁵Howard Hughes Medical Institute, Boulder, Colorado, USA; ⁶Biofrontiers Institute, University of Colorado, Boulder, Colorado, USA.

sequence fragments to organisms and functions is improving as more human gut microbial genomes are sequenced and annotated²², and as more complementary ‘omics’ data sets become available^{23,24}.

Despite the highly divergent compositions of gut microbiota across individuals, the functional gene profiles are quite similar (Fig. 4). This similarity was reported in a study of 18 females who shared more than 93% of the enzyme-level functional groups, but few genus-level phylotypes²⁰, and was confirmed in much larger populations in HMP and MetaHIT data^{13,14}.

The gut microbiota all contain pathways for central, carbohydrate and amino-acid metabolism²⁰. However, not all pathways are represented, and grouping genes into broad functional categories can conceal the finer differences in function between individuals. Variable functions that are restricted to a species or strain — including pathogenicity islands, vitamin and drug catabolism, motility and nutrient transporters — are possible targets for personalized diets and therapeutic strategies.

Many genes are expressed only under specific conditions. Shotgun sequencing approaches that measure levels of mRNA (metatranscriptomics) or shotgun proteomics (metaproteomics) may uncover functional variation with factors such as disease or diet that DNA studies overlook. For example, genes involved in carbohydrate metabolism and energy generation have been found to be expressed as proteins at higher levels than predicted from metagenomic data, underscoring the importance of these processes in the gut²³.

The observation that we share a functional core microbiome but not a core microbiota can be related to macroecosystems. For instance, rainforests that look the same and have many of the same functions can be composed of different species that have independently evolved. The challenge of ecological studies of the gut is to understand which members of the community have similar functional niches and which ones can be substituted for one another. Although taxonomic composition is far more variable than functional gene composition, at a general process level at least, several studies have shown a correlation between the two^{4,14,25}. Functional gene composition variation that is not well explained by taxonomic variation can indicate that phylogenetically unrelated taxa have converged on similar functions and may be crucial for identifying and understanding functional components associated with altered physiological states (Fig. 2).

Factors that drive normal variation

Having established that the healthy gut microbiota is highly variable, we must next understand why it varies, so that this information can be used to tailor therapies or clinical trials. For example, the extent to which members of the same family harbour similar microbes will determine whether the family history of microbiota-driven disease is informative. How much the microbiota varies with age or pregnancy should be taken into account when designing cohorts. The sensitivity of the microbiota to external factors, such as diet, will inform strategies for treating microbiome-linked diseases.

Age

Marked changes in the gut microbiota occur early on in life, with an increase in diversity and stability over the first three years^{4,26,27} (Fig. 5). The maturation of the human microbiota is an example of ecological succession^{4,26,27}. After the initial colonization, communities undergo consecutive changes in composition and function until a relatively stable climax community is established.

The microbiota in infants is relatively volatile. Variation in microbial communities and functional gene repertoires between individuals is greater in infants than in adults. This has been seen in rural communities in Malawi and Venezuela, and in US metropolitan populations of European and African descent⁴. However, infant microbiomes do share characteristic properties across individuals and populations, both compositional (many bifidobacteria, and

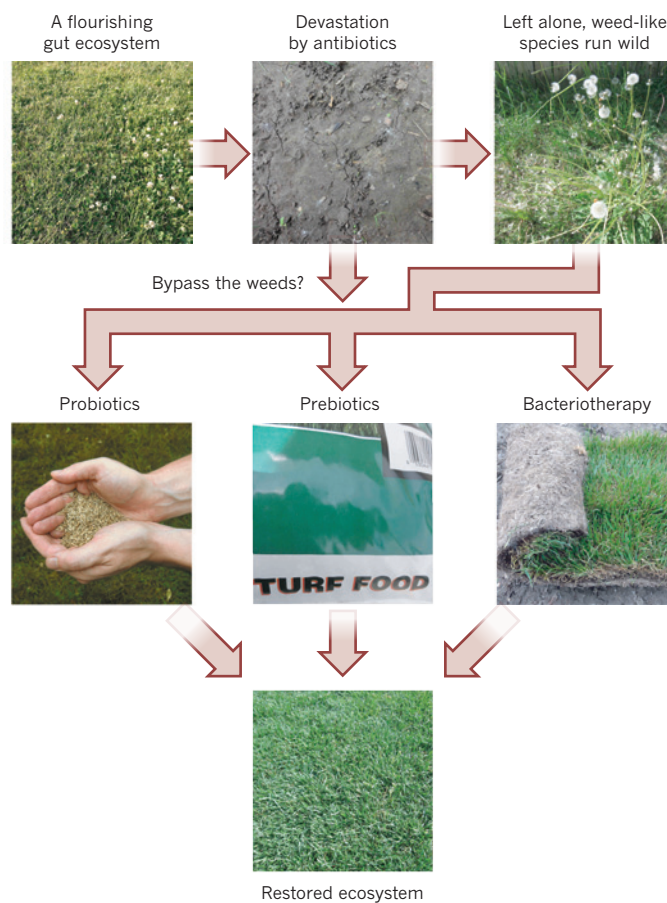


Figure 1 | Maintaining our gut microbial lawn. In some ways, maintaining a healthy microbiota is like lawn care: severe interventions can take the ecosystem back to bare earth, so it has to be re-established. Although many people recover naturally, this is not guaranteed, and weed-like species that are adapted to perturbed ecosystems often run wild. Several strategies can be used to restore the ecosystem: reseeding with a few well-defined ‘good’ microbes (probiotics), adding compounds that are thought to specifically promote the growth of beneficial microbes (prebiotics) and transplanting the entire microbial ecosystem, for example from a stool sample (faecal bacteriotherapy). The weed stage could be bypassed by administering the treatment immediately after the disturbance before the weeds can establish themselves. An additional strategy (not shown) is to use specific drugs, such as narrow-spectrum antibiotics, that target undesirable members of the microbial community. Although we are beginning to learn what a healthy microbial community looks like, and to recognize signs of weed species, our understanding of which strategies for altering the microbiota work best, and predicting which will work for a given individual, is still in its infancy.

lower species richness than in adults) and functional (more genes that encode enzymes involved in folate biosynthesis⁴).

The microbiota of infants is affected by antibiotic use, whether they were breastfed and the method of delivery, although whether differences in the microbiota in early life affect the composition of the adult microbiota is not well understood^{28,29}. However, differences in composition driven by factors in infancy may affect susceptibility to immunological diseases into adulthood, such as asthma and atopic diseases²⁹. A possible mechanism for this susceptibility has been demonstrated in germ-free mice that accumulate invariant natural-killer T (iNKT) cells in the colonic lamina propria and lungs. The mice were protected from the associated pathology of this accumulation if colonization with a conventional microbiota occurred in neonatal, but not adult germ-free mice, indicating that infancy is a crucial time for contact with the microbiota⁵. Identifying the components of the human microbiota that shape our immune system in early life is now needed.

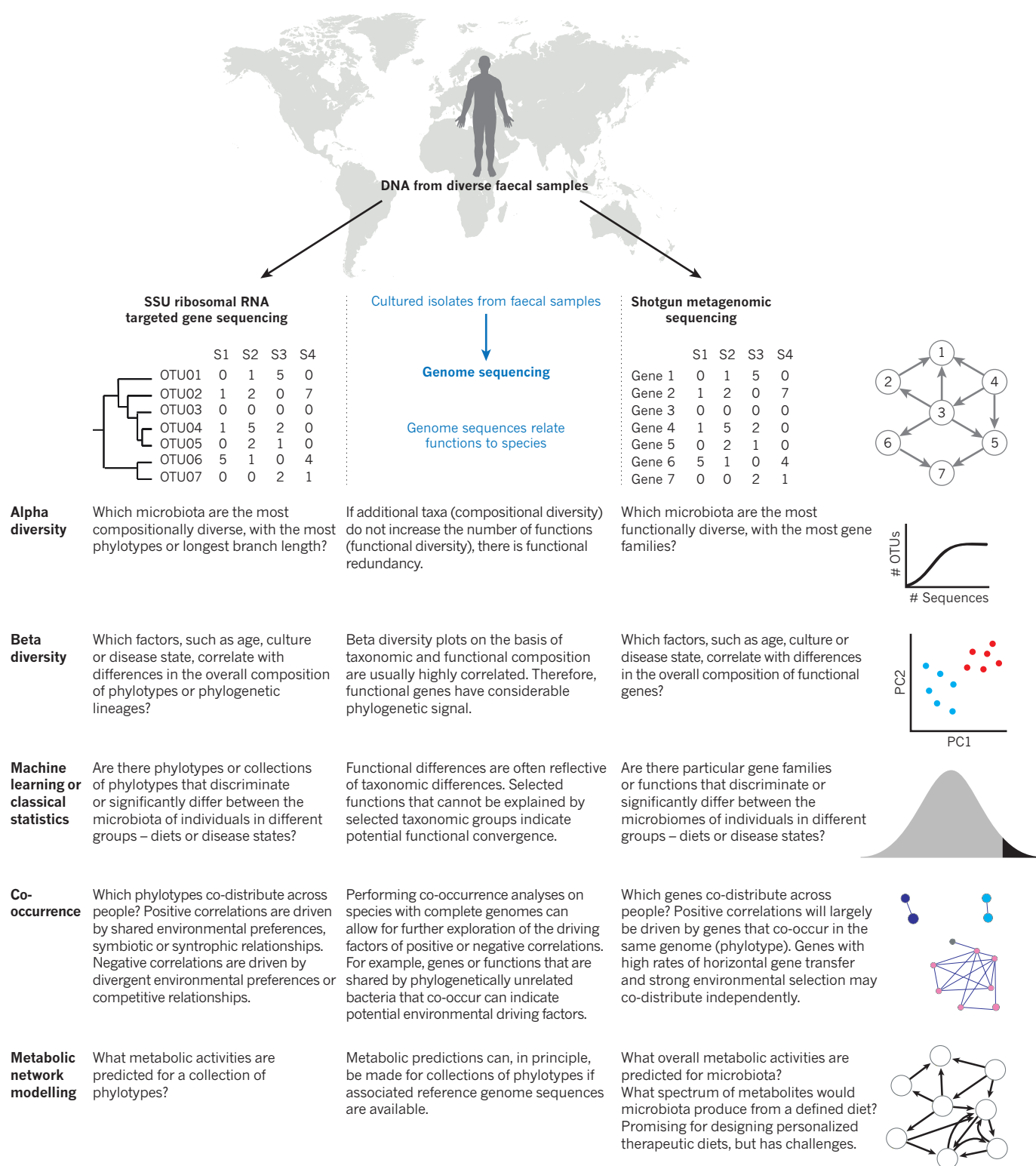


Figure 2 | Tools for evaluating microbiota diversity. Targeted sequencing of phylogenetically informative genes (usually small subunit (SSU) ribosomal RNA) or random sequencing of all genes (shotgun metagenomic sequencing) can be used to assess DNA extracted from faecal samples. Genome sequences from cultured isolates link these two data sets by indicating which species contain which genes, and therefore functions. Shotgun metagenomic data will become substantially more useful as the number of reference genomes continues to increase with additional strain sequencing efforts. SSU rRNA gene sequences are related to each other in the form of phylogenetic trees because related phylotypes (clusters of similar sequences defined by sequence similarity) generally have more similar functional attributes. Genes that

encode proteins that perform known enzymatic reactions are related using metabolic networks, because genes that participate in the same metabolic pathway can work together to produce a phenotype. Data from SSU rRNA targeted gene and shotgun metagenomic sequencing can be analyzed by the same suite of techniques (detailed in the table and with illustrations to the right). Combining analysis results from targeted and shotgun sequencing data from the same samples, and sometimes using genome sequences to relate the two, can provide additional insight (middle column). Combining targeted sequencing with mRNA, protein and metabolite level analyses will be crucial for moving from *in silico* predictions to direct measurements of expressed community properties. OTU, operational taxonomic unit.

Genetics, environment and diet

How factors such as the environment and diet shape the human gut microbiota is still unclear, partly because these factors are often confounded. Twins and mother-daughter pairs have more similar microbiota compositions than unrelated individuals, suggesting that there could be a genetic influence over the microbiota^{9,20}. However, monozygotic and dizygotic adult twins have equally similar microbiota, suggesting environment rather than genetics may drive familial similarities²⁰.

Populations can be separated by characteristic differences in the gut microbiota. For example, Italian children have a different microbiota from children from rural Africa³⁰, and both children and adults from the United States have a very different microbiota from populations in Malawi and the Amazonas state of Venezuela⁴ (Fig. 5). Although genetically different, these populations also differ in other factors that could affect the microbiota, such as environmental exposures, provision of adequate sanitation and levels of cleanliness, diet and antibiotic use. Cultural factors, especially diet, may be crucial in shaping the gut microbiota. The ratio of two of the main genera of gut bacteria, *Prevotella* and *Bacteroides*, correlates well with the overall pattern of diversity across healthy adults (see discussion of enterotypes later)^{31,32}. Levels of *Prevotella* were enriched in children from a rural African village who had a high-fibre diet³⁰, and in children and adults from Malawi and Venezuela whose diet was dominated by plant-derived polysaccharide foods such as maize and cassava. By contrast, the microbiota of those in the United States had more *Bacteroides*⁴. For healthy adults in the United States, differences in long-term diet between individuals also was correlated with these bacteria. *Bacteroides* was associated with a long-term diet rich in animal protein, several amino acids and saturated fats, and *Prevotella* was associated with carbohydrates and simple sugars³². The relative importance of changes in the microbiota as a result of diet, compared with other factors that affect health, remains a subject of active investigation.

Understanding how cultural traditions affect the microbiota should highlight the factors that result in the marked differences in the incidence of diseases associated with microbiota. For example, the incidence of IBD and other allergies is greater in industrialized, Western, societies than in agrarian cultures³³. Understanding these associations will require expanded studies that sample a greater number of populations and control the confounding factors. This may come in the form of a human microbiome diversity project that parallels the Human Genome Diversity Project³⁴. Especially valuable will be populations in which variables that correlate with microbiota diversity — such as history of antibiotic use, diet and environmental exposures — can be measured. Comparisons with environmental exposures may require sampling from the environment as well as from individuals. Because exposure to a microbiota in early life is particularly relevant to the development of immunological disease⁵, prospective studies that include infants or even prenatal studies, will be especially valuable. When determining disease aetiology, immigrant populations can provide an opportunity to disentangle host genetics, geography and culture because the incidence of some diseases matches that present at the destination rather than at the origin³⁵.

Microbiota variability, and human health and medicine

Understanding the compositional and functional differences in the gut microbiota can lay the foundation to relate these differences to human health. Differences in the microbiota and the microbiome could help to explain the variation in the gut metabolic processes of individuals, including the metabolism of drugs and food^{11,36}. Many of these metabolic pathways are outside the common functional core, so they can underlie host-specific responses. For example, the health benefits of diets rich in soya, such as improvements in vasomotor symptoms, osteoporosis, prostate cancer and cardiovascular disease, have been attributed to (S)-equol produced from the soya isoflavone daidzein by bacterial rather than human enzymes³⁷. Only 25–30% of

the adult population of Western countries produce (S)-equol with a diet of soya-rich foods, compared with 50–60% of adults from Japan, Korea or China³⁸. The cancer-protective effects of soya beans that have been described in Asian populations may therefore not be true of Western populations. Similarly, the gut microbiota can indirectly determine whether the widely used analgesic paracetamol will be metabolized to paracetamol sulphate or paracetamol glucuronide, potentially altering its efficacy and toxicity³⁶. The microbes mediate this metabolic phenotype by producing the compound *p*-cresol, which competes with paracetamol for human enzymes that catalyse sulphonation³⁶. Substantial differences in the gut microbiota between populations and concomitant effects on drug metabolism will necessitate population-specific drug trials for examining effectiveness and toxicity. Understanding how the microbiota varies across the population, and correlating this variability with specific microbial functions, is emerging as a component of personalized medicine.

Stable configurations of healthy microbiota

The landscape of stable states for the human gut is still unknown. Samples obtained over time from the same individual are more similar to one another than those obtained from different individuals, suggesting that each person has a relatively distinct, stable community^{20,39–42}. This temporal stability could be related to the concept that stable equilibrium states exist for the microbiota, for which disturbances, stochasticity and temporal dynamics of individual microbes produce change, but the community is still drawn to a central attractor^{43,44}.

Clinically, it would be convenient if a few clearly differentiable stable states could be used to stratify the gut microbiome across the human population. For example, if microbes in one state were known to metabolize a drug into harmful metabolites, a simple test could avoid complications for the patient, as in the case of paracetamol and liver toxicity³⁶. It would be especially helpful if the same states could be applied across the board so that, for example, the same community features that correlated with the generation of toxic metabolites from a drug also correlated with obesity.

Enterotypes propose this concept. The human microbiome has been described to form three distinct host–microbial symbiotic states driven by groups of co-occurring species and genera, which are characterized by a relatively high representation of *Bacteroides*, *Prevotella* or *Ruminococcus*³¹. Although the overall clustering structure, which was based on 33 faecal shotgun metagenomes, was not statistically significant, the three enterotype clusters described the real data better than randomly generated data sets³¹. Similar patterns were reported after existing 16S rRNA data from 154 individuals from the United States²⁰ and metagenomic data from 85 individuals from Denmark¹³ were reanalysed. However, subsequent studies of 98 healthy adults from the United States³², 531 healthy infants, children and adults from Malawi, Venezuela and the United States⁴, and 250 healthy adults from the United States¹⁴ failed to show the same pattern. Variation among adults in these populations was, however, associated with a trade-off between *Prevotella* and *Bacteroides*, suggesting an important role for these taxa, or taxa that co-distribute with them, in structuring the microbiota. When infants were included in the analysis, variation was also associated with a trade-off between the infant-associated *Bifidobacterium* genus and lineages that were common only to adults⁴ (Fig. 5). The number of unique configurations that form functional, stable communities may be large and may not be easy to classify into a manageable number of distinct 'types'. Considering this concept further in more extensive data sets made up of different ages, cultural traditions or geographic locations, and physiological or disease states will be important.

Differences in the microbiota with disease

The composition of the microbiota differs with the physiological state of the host. People who are obese harbour fewer types of microbes in their gut than people who are lean, and have a significantly different abundance of specific taxa and functional genes^{6,7,45}. Classification

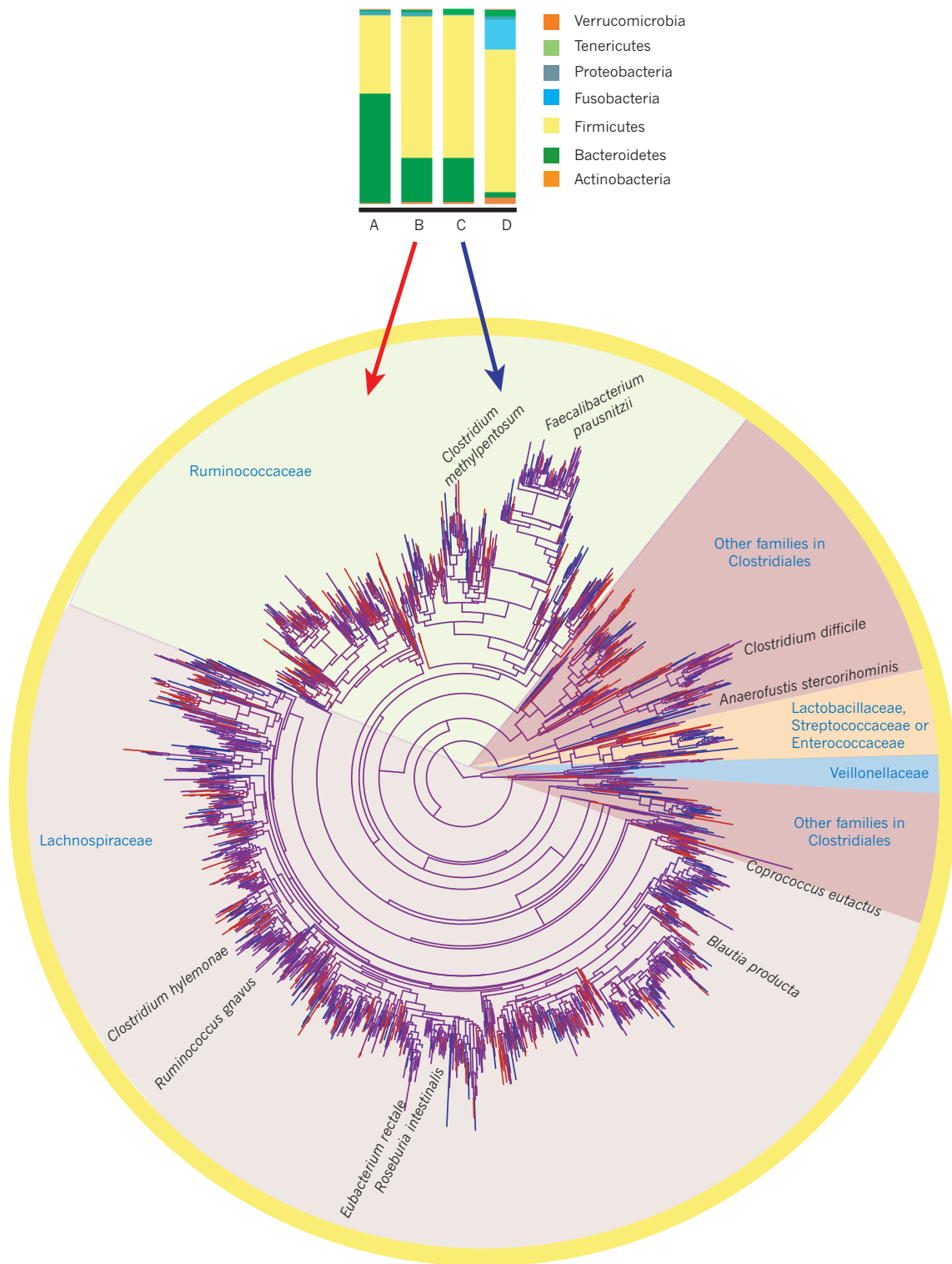


Figure 3 | Diversity of the human microbiota at different phylogenetic scales. The human microbiota is remarkably varied within and between individual hosts. Although this complexity can be simplified by evaluating communities at higher taxonomic levels, the species within each phylum have different biological properties, and significant changes detected at higher taxonomic levels are probably driven by only a subset of the species in those higher taxa. Using the 16S rRNA sequence data of four of the adults from the United States in ref. 4, the high diversity and variability among individuals, and the degree to which taxonomic grouping at high levels can mask this diversity is shown. Phylum level diversity can have a marked variation even across healthy adults in the same population.

Individual A has an unusually high proportion of Bacteroidetes; individual D has an unusually high proportion of Fusobacteria; and individuals B and C are dominated by Firmicutes and Bacteroidetes, which is the more typical phylum-level distribution for this cohort. However, even B and C differ at finer scales. The tree depicts the phylogenetic relationships between species-level phylotypes in just the most diverse of the phyla, Firmicutes, in individuals B and C. Branches specific to individual B are red; branches specific to individual C are blue; and shared branches are purple. Each individual has many unique phylotypes not found in the other. As described in many surveys of the human gut^{14,18,40}, the Ruminococcaceae and Lachnospiraceae families are particularly rich in phylotypes.

of people as lean or obese can be made solely on the basis of their gut microbiota with 90% accuracy^{46,47}, but they do not separate into distinct microbiota-based clusters on commonly used principal coordinates plots, which are used to identify statistical differences between groups. Thus, multiple statistical techniques are needed to show fully the differences in the microbiota between different physiological states (Fig. 2).

Some differences in the microbiota can contribute directly to disease states. Gnotobiotic mice that were raised germ-free then colonized with the microbiota from an obese mouse gained fat more rapidly than those colonized with the microbiota of a lean mouse^{7,45}. A phenotype can emerge from different compositional backgrounds, which may indicate that specific components of the microbiota can exert large effects or that many different changes can lead to the same functional result.

Differences in faecal microbial community diversity, composition and function have also been correlated with Crohn's disease⁹, ulcerative colitis¹⁰, irritable bowel syndrome (IBS)⁴⁸, *Clostridium difficile*-associated disease (CDAD)⁴⁹ and acute diarrhoea⁵⁰. Sometimes, the nature of the microbiota deviation from health is consistent across individuals with the same disease. For instance, a twin study of IBD found marked and reproducible deviations in patients with ileal Crohn's disease relative to the controls, and more subtle, but characteristic, changes in patients with colonic Crohn's disease⁵¹, and specific functional differences were also observed from metabolic profiling of the same samples²⁴. Other diseases are associated with marked deviations from health that are inconsistent across individuals. For instance, individuals with recurrent CDAD had a phylum-level diversity that was very different from controls but not similar to each other⁴⁹. Many disease studies are confounded by extensive use of treatments, such as antibiotics, that may obscure true disease-associated changes, highlighting the urgent need for prospective longitudinal studies that establish cause and effect.

Parallels between host physiological states

Studies of the microbiota often target one specific disease or state, but comparisons of the microbiota across many diseases can show common changes in the gut environment. Disturbed mucous layers that line the intestinal cell wall and concomitant inflammation are seen in individuals with IBD, coeliac disease, HIV enteropathy, acute diarrhoea, diverticulosis, carcinoma and IBS⁵². Given these parallels, an increase or decrease in abundance of similar microbes across different disturbances might be expected⁵³, but elucidation of these differences may require detailed biogeographical studies along the length of the gut — once safe and reliable means for such comprehensive sampling are developed.

Perturbed adult gut microbial communities are intriguingly similar to infant gut microbial communities. Both systems may represent successional communities in which the same opportunistic or fast-growing species can predominate⁵³. For instance, *C. difficile* is a normal gut resident that can cause disease when antibiotics compromise the stable adult gut communities; it also colonizes 2–65% of infants, although most infants are asymptomatic^{54,55}. *Clostridium bolteae* and *Clostridium symbiosum* are also associated with a disturbed gut and systemic infection, and are found in the infant gut⁵³. The microbiota of individuals with ileal Crohn's disease can also resemble that of infants: both have increased levels of *Ruminococcus gnavus* and Enterobacteriaceae in their stools, and an under-representation of the genera that are prevalent in healthy adults, including *Faecalibacterium* and *Roseburia*⁵¹. These examples show the importance of understanding whether generally opportunistic members of the gut microbiota have a selective advantage during early succession or disruption caused by disease, and therefore whether they are the side effects of disease rather than causal agents.

Resilience of stable states

Resilience is the amount of stress or perturbation that a system can tolerate before its trajectory changes towards a different equilibrium state⁵⁶. Macroecosystem studies of human interference such as

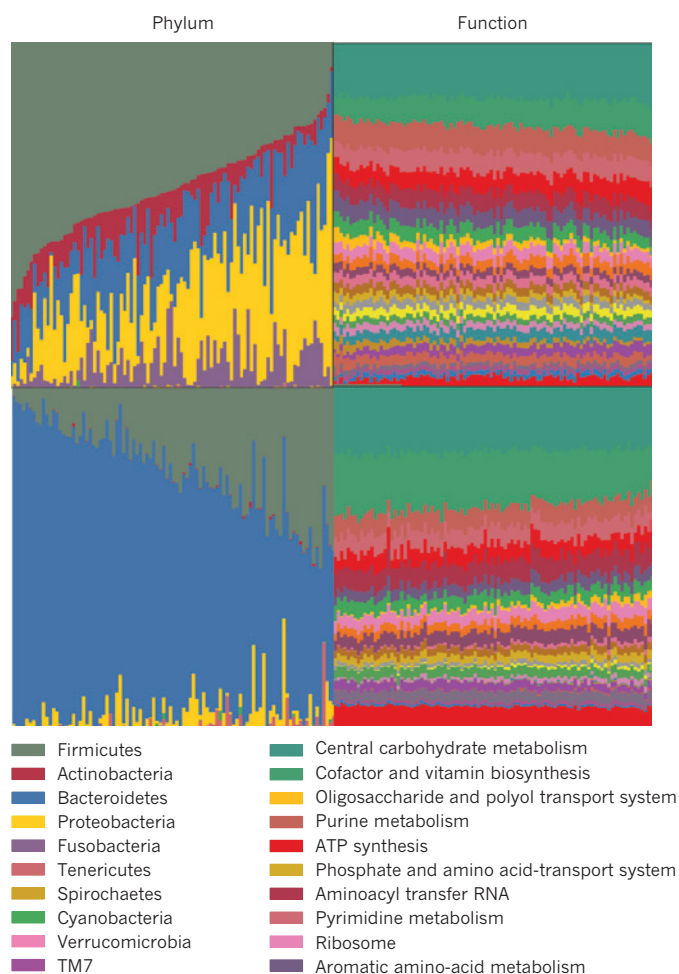


Figure 4 | Functional redundancy. The functional redundancy in microbial ecosystems may mirror that in macroecosystems. As shown in the HMP data set¹⁴, oral communities (top panels) and faecal communities (bottom panels) analysed using 16S rRNA (coloured by microbial phyla, left panels) show tremendous abundance diversity. The same samples analysed by shotgun metagenomics (panels on right) have remarkably similar functional profiles. Reprinted with permission from ref. 14.

resource exploitation, pollution, land-use change and global warming have shown how ecological systems can be transformed into less productive or less desirable states⁵⁶. For example, in a tropical lake community, a regime in which submerged plants dominate is preferred to one with extensive free-floating plant cover because dense mats of floating plants create anoxic conditions that reduce animal biomass and diversity. However, pollution can cause floating plants to predominate because they are better at competing for light and can exclude submerged plants when the nutrient load is high. By understanding the environmental drivers of conversion between states, interventions can be used to induce a regime change. For example, harvesting the floating plants once can induce a permanent shift to the submerged-plant-dominant state, but only if the nutrient loading is not too high⁵⁷.

Environmental studies can provide examples of the microbial response to perturbations, and perhaps an insight into how the gut microbiota might react. During the Deepwater Horizon oil spill in the Gulf of Mexico, the microbial community structure and functional gene repertoire in the deep-sea oil plume shifted, transiently, to an enrichment of microbes that were capable of hydrocarbon-degradation^{58,59}. This is similar to the effect of an extreme dietary change on the gut microbiota of mice that were moved from a low-fat, plant-rich diet to a high-fat, high-sugar diet⁷. In both cases, the

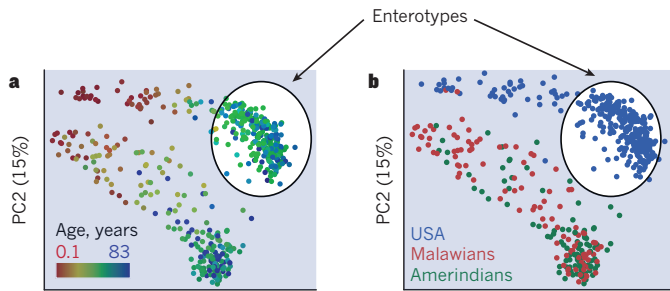


Figure 5 | Human microbial diversity and enterotypes. Enterotypes³¹ were determined when evaluating only adults from the United States and Europe (circled in white). By including children from the United States and children and adults from developing countries, the picture of human-associated microbiota diversity greatly expands. The relationship between the microbiota of 531 healthy children and adults from Malawi, Amazonas state of Venezuela (Amerindians) and the United States was evaluated using sequences from the 16S rRNA gene in faecal samples and a principle coordinate analysis of unweighted UniFrac distances (adapted with permission from ref. 4). **a**, Infants differentiate strongly from adults, and **b**, adults from the United States have a distinct composition from those of Malawi and Venezuela, indicating the diversity differences are mainly owing to age and culture.

microbial communities shifted substantially, probably because of the exposure to new substrates that provided a selective advantage to certain members of the community. Whether a particular disturbance will disrupt a stable state depends on the resilience of the community to a type of perturbation.

Dietary changes

Understanding the resilience of the gut microbiota is crucial for determining the efficacy of therapeutic diets. Consuming a carbohydrate- or fat-restricted low-calorie diet for a year⁶, or a high-fat and a low-fibre, or low-fat and high-fibre diet for 10 days³² can induce statistically significant changes in the gut microbiota. However, these changes in species and gene content are small compared with the baseline variations that occur between individuals. Long-term dietary surveys and cross-cultural comparisons suggest that changes to diet might lead to regime changes over longer periods of time^{4,32}, perhaps by eroding the landscape of alternative stable states to allow changes that short-term ‘nudges’ cannot.

Antibiotics

Antibiotic administration can move the microbiota to alternative stable states. In healthy volunteers who were given two courses of ciprofloxacin in a 10-month period, the faecal microbiota reached a stable state similar to, yet distinct from, the pretreatment state⁴¹. However, the magnitude of disturbance that occurred after ciprofloxacin treatment, and the speed and extent of recovery to the pre-ciprofloxacin state, suggested that resilience of the microbiota varies across individuals and between ciprofloxacin treatments in the same individual⁴¹.

Long-term studies of the microbiota after antibiotic treatment indicate that post-antibiotic equilibrium states are themselves resilient. For example, clindamycin treatment affected *Bacteroides* in the gut for up to 2 years after treatment had finished⁶⁰. Similarly, three individuals with dyspepsia treated for 1 week with a combination of metronidazole, clarithromycin and omeprazole had a shift in their microbiota state that persisted for up to 4 years without additional antibiotic treatment⁴². In both cases, a significant increase in the antibiotic-resistance genes persisted for years^{42,60}, suggesting that the post-disturbance state would probably be more resilient to the same disturbance again because a greater proportion of the microbiota would be resistant. This is consistent with a case from the study of healthy volunteers given the two courses of ciprofloxacin⁴¹. Initial recovery of the microbiota was slow and incomplete, stabilizing in a different interim state, but after a second treatment, recovery was relatively quick. But in the same study, a different case had the opposite

response, in which an essentially complete recovery occurred after the first ciprofloxacin treatment, but stabilized to a distinct state after the second treatment, suggesting that the initial antibiotic treatment decreased resilience⁴¹, and that the resilience of the microbiota to future antibiotic treatments can vary considerably between individuals.

New species invasion

Resilience of the microbiota to challenge with exogenous microbes is also important. The gut microbiota generally shows colonization resistance, in which the native microbiota prohibits harmful (pathogenic)^{1,2,61} and potentially beneficial (probiotic)⁶² microbes from establishing. Challenging the gut microbiota with a foreign microbiota rather than with specific pathogens or probiotics has suggested the native community may be less resilient to colonization by exogenous microbes than previously thought. When 14 conventionally raised Lewis rats were force fed with pooled microbiota from Sprague-Dawley and Wistar strain rats, the microbiota diversity of the Lewis rats increased substantially and changed to resemble that of the Sprague-Dawley and Wistar rats, and phylotypes established from the donors persisted for up to 3 months after transplantation⁶³. In contrast to ecological theory — that disturbances eliminating native species facilitate the establishment of exotic species⁶⁴ — transplantation after the resident microbiota have been depleted with two broad-spectrum antibiotics reduced rather than promoted the establishment of the donor microbiota⁶³.

The success of microbiota transplantation in the treatment of recurrent CDAD provides further support of the plasticity of an established gut microbiota when challenged with a complex one. Bacteriotherapy⁶⁵ involves introducing a faecal sample from a healthy donor as a homogenate by injection into the caecum using a colonoscope. In one case study, the symptoms of a patient with recurrent CDAD were resolved after faecal transplantation from a healthy donor. Before treatment, the recipient had a disturbed microbiota with *Veillonella* and *Streptococcus* predominating, whereas the donor microbiota was dominated by *Bacteroides*. A month after treatment, the patient had no *C. difficile* in stool samples, but the donor species were present, indicating that the donor microbiota persisted over this interval. Faecal transplantation for CDAD is becoming increasingly common and seems to be highly effective. A survey of 317 patients across 27 case series reported disease resolution in 92% of cases, with a single treatment being sufficient in 89% of cases⁶⁶. The success of bacteriotherapy suggests that a previous course of antibiotic treatment reduces the resident microbiota to the extent that a donor microbiota can establish itself. However, because faecal transplantation is not a first-line therapy, there are no data on its effectiveness without antibiotics (A. Khoruts, personal communication). Understanding the successional processes and which microbial taxa allow a healthy microbiota configuration to be established in faecal transplant recipients will require further studies. These studies will allow the development of a standardized way of formulating and analysing the donor specimen, design and interpretation of dosing studies, and analysis of short- and long-term safety.

Mechanisms for resilience

Understanding how the microbiota forms a stable, resilient state would allow strategies to be devised to increase the resilience of healthy states or decrease the resilience of unhealthy states. For example, a healthy state with a high resilience to exogenous microbes might mean one person at a dinner party escapes food poisoning but their companions with lower resilience fall ill. A degraded microbiota with high resilience can contribute to chronic diarrhoea or inflammation, as is inherent with CDAD, IBD and IBS.

Species and functional response diversity

The aspects of diversity that are crucial for conferring resilience in macroecosystems are probably the same features that are important in microbial ecosystems. One such parameter is the number of species present in a given system, or species richness. (In

culture-independent studies, this number depends on the number of sequences collected per sample.) Species-rich communities are less susceptible to invasion because they use limiting resources more efficiently, with different species specialized to each potentially limiting resource⁶⁴. An excess of nutrients in water, or eutrophication, often decreases ecosystem diversity because a small number of species overgrow and outcompete everything else, with a concomitant decrease in resilience⁶⁷. In a similar way, decreased diversity has been linked to obesity and to a diet that is high in fat and sugar compared with one that is low in fat and plant-based^{4,7}. Whether this decrease in microbiota diversity results in a decrease in resilience is not known. Low microbiota diversity also correlates with IBD⁵¹ and recurrent CDAD⁴⁹, but the effects on microbiota resilience are not known.

Functional response diversity is the degree to which species in a community that contribute to the same ecosystem function vary in their sensitivity to ecosystem changes⁶⁸. High functional response diversity may, for instance, allow a species that is relatively rare but functionally similar to fill a niche when an abundant species is compromised by an environmental disturbance⁶⁸. In a macroecology example, a compromised coral reef that was in a healthy state changed to an unhealthy, algal-dominated state only when both the algal-grazing fish were overfished and the sea urchins that had increased in numbers to fill this niche were compromised by a pathogen⁶⁸. The same principle is also likely to apply to the gut microbiota. Following antibiotic treatment, a previously rare microbe may increase in abundance to fill a niche that had been dominated by a microbe with higher antibiotic sensitivity, leading to persistence of the same stable state but with a decreased resilience because of the decrease in functional redundancy.

Human-gut-adapted bacteria are likely to have high functional response diversity because phylogenetically disparate microbes often perform similar metabolic functions. For example, methanogenic Archaea, sulphate-reducing bacteria and phylogenetically diverse acetogens in humans and mice all consume hydrogen generated by other microbes during fermentation⁶⁹. Butyrate producers of the Clostridiales order can have different ecological strategies, such as adaptation to different stages of community succession, oxygen tolerance and substrate preference. The abundance of *Anaerostipes caccae* peaks in infancy, whereas *Eubacterium hallii* and *R. intestinalis* are more abundant in adults. *A. caccae* is able to survive 10–60-minute periods of exposure to air better than *E. hallii* or *R. intestinalis*⁷⁰. Furthermore, *E. hallii* but not *F. prausnitzii* can use lactate as a substrate⁷¹. Thus, butyrate production in the gut can also continue through different successional and metabolic states.

Competition and feedback loops

The densely populated gut environment means that microbes compete to use the same resources or inhibit each other directly using antimicrobial products. Phylogenetically related bacteria could be expected to compete because of their overlapping functional roles, habitats or both, yet contrary to this expectation, phylogenetically similar species tend to appear in the same samples⁷². For example, analysis of bacteria with complete or draft genome sequences from 124 individuals from Europe found that the abundance of related Enterobacter species, including *E. coli*, *Salmonella enterica*, *Citrobacter koseri* and *Enterobacter cancerogenus* were positively correlated across individuals⁵³. These related species may share environmental preferences so that they are selected for simultaneously. The abundance of closely related species can also predict the susceptibility to intestinal colonization by both pathogenic and commensal bacteria in mice⁷³.

Feedback loops can either stabilize or destabilize the microbiota (Fig. 6). Stable physiological states are preserved by negative feedback, in which a change to the gut environment results in opposing changes that maintain homeostasis. This feedback is likely to be controlled by a tight interplay between microbial metabolic activities

and host pathways. For example, microbial metabolites might induce changes in the expression of the host pathways that control gut retention time so that it rises above or below the optimum, causing diarrhoea or constipation. This deviation probably induces host signalling pathways to correct it. This is similar to body temperature regulation, in which a rise above the optimum induces thermoregulatory mechanisms such as sweating, and a fall induces mechanisms such as shivering. Stabilizing the physiological parameters that are controlled by negative feedback from the host could therefore promote resilience of the microbiota.

Negative-feedback loops that promote microbiota resilience could also operate independently of the host. For example, if the abundance of a particular microbe exceeds a certain threshold, this would result in a change in the gut environment that would decrease the growth of that microbe relative to other species. These feedback loops may involve the accumulation of a phage specific to that microbe or the accumulation of a specific toxic metabolite. Negative feedback is thought to promote high ecosystem diversity, which can promote resilience⁷⁴.

Positive feedback is thought to induce an ecosystem change because a difference from a point in one direction produces further change in the same direction. However, positive-feedback mechanisms could support stability at the individual microbe level or of the microbial consortia that promote each other's growth. Metabolic activities or interactions with host pathways in microbes may induce a physiological state that favours their growth over the growth of potential competitors, promoting resilience. The microbiota is likely to contain both important functional drivers of physiological status and microbes that co-occur with these drivers because they are able to thrive in such an environment. Invasion by microbes that do not thrive in that particular physiological state would be prevented, for example if a microbiota containing functional drivers of a low inflammatory state resists colonization by pathogens, and microbiota with functional drivers of a relatively high inflammatory state resists colonization of beneficial microbes (Fig. 6).

Positive- and negative-feedback loops are also likely to have a role in destabilizing the microbiota during regime changes, such as during succession in early development and following a disturbance (Fig. 6). Negative feedback, whereby an organism's activity alters the environment so that fitness is decreased, can induce a directional change when microbes induce a physiological state that favours their competitors. For instance, the higher redox potential in the gut of infants is probably one of the factors that explains the relative success of facultative anaerobes such as *E. coli* or some *Lactobacillus* in early development⁷⁵, but the reduction of oxygen that results from their metabolism favours their eventual replacement by a consortium dominated by strict anaerobes.

The importance of feedback for successional changes is also likely to involve a complex interplay between the microbiota and its host. For example, the same change in redox potential that directly affects microbial fitness can also affect the expression of host factors in the gut epithelium, such as hypoxia-inducible factor, in early development and inflammatory diseases of the gut⁷⁶. Manipulating positive and negative feedback at the level of the host, the individual microbes and the entire gut ecosystem will be essential to maintaining healthy stable states and switching from an unhealthy to a healthy state.

Prospects

High-throughput sequencing of samples suggests that the microbiota of each person has some resistance to perturbation, but that this resistance can be overcome by diet, drugs, prebiotics or probiotics. Dietary changes may alter the regime in the gut over long time periods. The surprising success of whole community transplants in healthy rats and in humans with CDAD shows that exogenous microbes can colonize the gut even with resistance from an entrenched microbiota. However, which microbes will be the best colonizers is unknown; neither do we know how a particular microbial

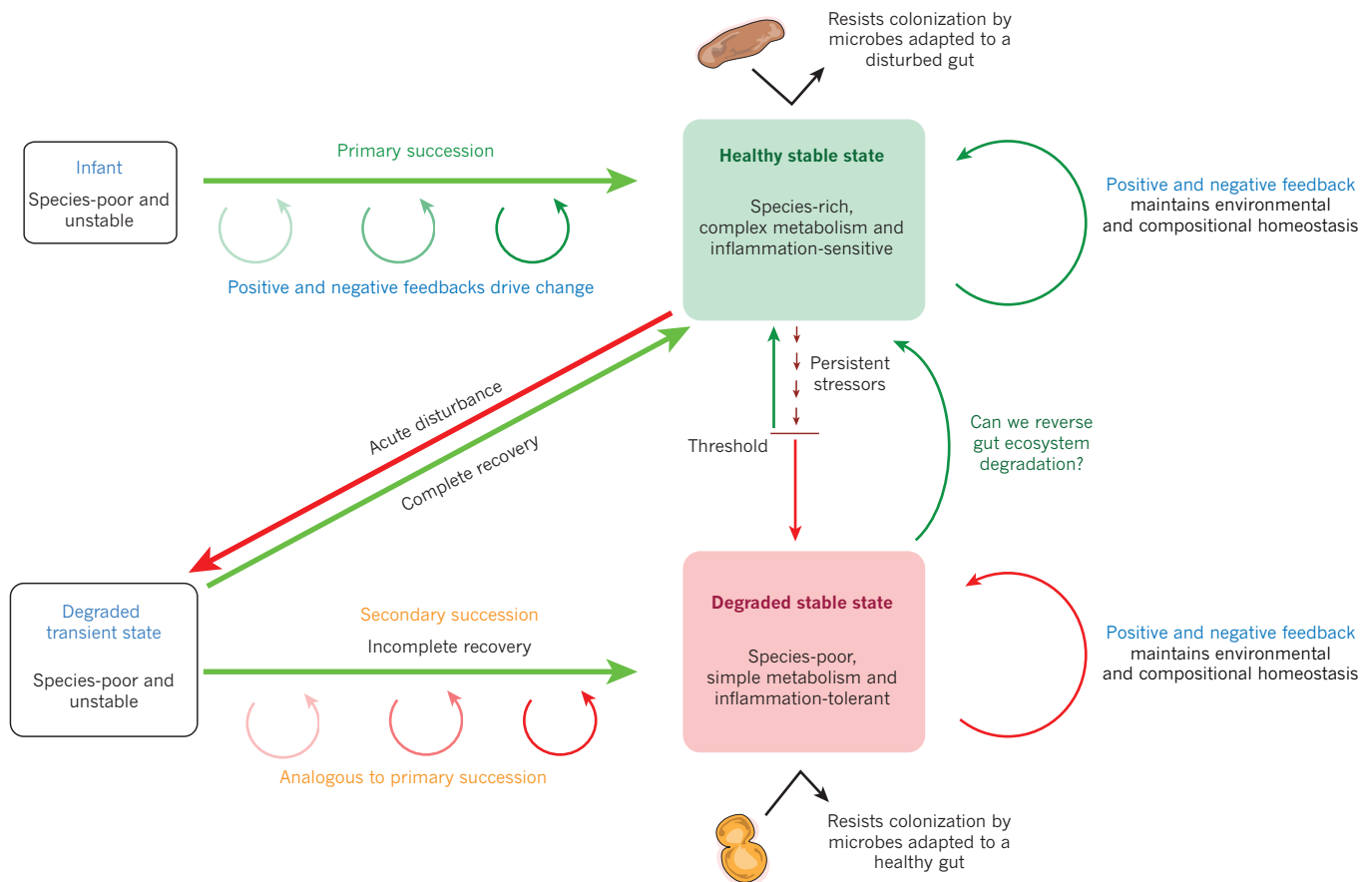


Figure 6 | Compositional transitions in the human gut microbiota.

Primary succession of the gut microbiota during early development involves a systematic turnover of species until a stable adult state is reached. Positive- and negative-feedback loops probably have a role in driving primary succession and in conferring resilience to healthy stable equilibrium states. Acute disturbances, such as antibiotic administration, are often followed by an unstable state that progresses to a stable state through secondary succession. In some cases, a complete recovery occurs, in which the stable state highly resembles the pre-disturbance state, but sometimes the post-recovery stable state is distinct. Post-disturbance stable states may be both degraded and resilient; for example, IBS that forms after an initial acute disturbance of the microbiota from an enteropathogen can persist in some individuals for years and even decades⁷⁸. Resilience

configuration and its functional attributes change in response to dietary components or exogenous microbes. We need to learn what conditions promote the health of desired species and exclude the undesirable ones, just as a gardener would exclude weeds.

The landscape of stable states for the microbiota and its implications for resilience is an important research direction. Small perturbations, such as short-term dietary changes, may allow the microbiota to return to its original state, but larger disturbances, such as antibiotic administration, could cause movement into a different state. The long-term implications for health are not yet well understood. Furthermore, changes to the stable equilibrium of the gut microbiota through long-term changes, such as inflammation, diet or repeated antibiotic administration, could make new states reachable with smaller perturbations. Host genetics, the factors that affect the microbiota early in development, diet and long-term drug administration may all contribute to individual differences. Consequently, the general landscape and the current community state could be important for determining individual responses to an intervention.

How resilient the microbiota is to diet, disturbance by antibiotics and challenge by exogenous microbes will have important implications for health care. The degree to which repeated applications of

of degraded states is probably driven by unique positive- and negative-feedback loops that occur both in concert with and independently of the host. Degradation to a stable state may also occur as a result of persistent stressors, such as poor diet, that slowly degrade resilience of a healthy state until a threshold is passed so that new feedback mechanisms are needed to maintain community composition and stability. Developing therapies that encourage transition from degraded to healthy stable states, or complete recovery to a healthy stable state following disturbance, may involve identifying species or combinations of species, and drivers of feedback loops. Whether interventions are more effective early in succession, when communities are more unstable but possibly stochastic, or later in succession, when convergence to the end point is more certain but the trajectory may be more difficult to change, will be important to investigate.

broad-spectrum antibiotics degrade the microbiota and its ability to function as a healthy system would need to be studied, especially in children, in whom the development of interactions between the microbiota and host is crucial⁵. However, identifying suitable controls is challenging given the large variation in and between individuals during early development. Regime change is not always instigated by acute disturbances and can occur gradually, as indicated by studies of resilience to dietary changes. Individuals whose microbiota has been degraded by long-term consumption of a high-fat and high-sugar diet may need long-term dietary changes to restore their microbiota to a healthy state.

The lower taxonomic diversity in individuals from Western cultures who have a high-fat and high-sugar diet raises the concern that global trends in diet could result in important microbial symbionts being lost from the broader population, possibly leading to the extinction of bacterial species that can provide important health benefits. Maintaining a collection of cultures from individuals in the developing world, and specifically from agrarian cultures, could help to preserve potentially important components of the microbiota.

Microbes often produce unique states in the gut through collective activities and cooperative metabolism; understanding associations

between disease states and sets of species rather than single taxa will therefore be important⁵³. A central problem of culture-independent metagenomic analyses is that identified phylotypes, or collections of these, often represent species for which there is little biological knowledge. Gut microbial ecology has come full circle, with the development of methods for culturing most of the diversity present in a community being given increasing attention so that the contribution that community members make can be further explored. The human faecal microbiota is largely composed of readily cultured bacteria⁷⁷. *In vitro* experiments with these isolates will be extremely valuable for exploring the metabolic attributes of a particular microbe or set of microbes, and the genes involved. However, because microbiota composition and function are also controlled by feedback from the host, *in vivo* studies in gnotobiotic mice will be particularly valuable. Indeed, the component of the microbiota that can be cultured has similar colonization dynamics, biogeographical distribution and response to dietary perturbations as the full microbiota, when transplanted into gnotobiotic mice⁷⁷. A translational science pipeline is developing in which particular phylotypes and their interactions that are important for health or that contribute to disease can be identified on the basis of distribution patterns with disease. Furthermore, the biological attribute that may be driving these patterns can be predicted from the expression or prevalence of functional genes in whole communities and in genomes. Cultures and animal models can be used to test and verify these hypotheses before human application. The tools are now in place and although the path will be difficult, the direction is clearer. ■

1. Candela, M. *et al.* Interaction of probiotic *Lactobacillus* and *Bifidobacterium* strains with human intestinal epithelial cells: adhesion properties, competition against enteropathogens and modulation of IL-8 production. *Int. J. Food Microbiol.* **125**, 286–292 (2008).
 2. Fukuda, S. *et al.* *Bifidobacteria* can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543–547 (2011).
 3. Sonnenburg, J. L. *et al.* Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005).
 4. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
 5. Olszak, T. *et al.* Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336**, 489–493 (2012).
 6. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
 7. Turnbaugh, P. J., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
 8. Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L. & Gordon, J. I. Human nutrition, the gut microbiome and the immune system. *Nature* **474**, 327–336 (2011).
 9. Dicksved, J. *et al.* Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J.* **2**, 716–727 (2008).
 10. Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl Acad. Sci. USA* **104**, 13780–13785 (2007).
 11. Gonzalez, A. *et al.* The mind–body–microbial continuum. *Dialogues Clin. Neurosci.* **13**, 55–62 (2011).
 12. Lupton, J. R. Microbial degradation products influence colon cancer risk: the butyrate controversy. *J. Nutr.* **134**, 479–482 (2004).
 13. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
 14. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
 15. Borenstein, E., Kupiec, M., Feldman, M. W. & Rupp, E. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proc. Natl Acad. Sci. USA* **105**, 14482–14487 (2008).
 16. Freilich, S. *et al.* Metabolic-network-driven analysis of bacterial ecological strategies. *Genome Biol.* **10**, R61 (2009).
 17. Claesson, M. J. *et al.* Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. *PLoS ONE* **4**, e6669 (2009).
 18. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).
 19. Reyes, A. *et al.* Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
 20. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009).
 21. Biagi, E. *et al.* Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **5**, e10667 (2010).
 22. Nelson, K. E. *et al.* A catalog of reference genomes from the human microbiome. *Science* **328**, 994–999 (2010).
 23. Verberkmoes, N. C. *et al.* Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* **3**, 179–189 (2009).
 24. Jansson, J. *et al.* Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS ONE* **4**, e6386 (2009).
 25. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).
 26. Koenig, J. E. *et al.* Succession of microbial consortia in the developing infant gut microbiome. *Proc. Natl Acad. Sci. USA* **108**, 4578–4585 (2011).
 27. Palmer, C., Bik, E. M., DiGiulio, D. B., Relman, D. A. & Brown, P. O. Development of the human infant intestinal microbiota. *PLoS Biol.* **5**, e177 (2007).
 28. Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl Acad. Sci. USA* **107**, 11971–11975 (2010).
 29. Kozyskyj, A. L., Bahreinian, S. & Azad, M. B. Early life exposures: impact on asthma and allergic disease. *Curr. Opin. Allergy Clin. Immunol.* **11**, 400–406 (2011).
 30. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107**, 14691–14696 (2010).
 31. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
- This paper reports that there is an association between co-occurring microbial groups, and that high *Prevotella* versus *Bacteroides* genus level abundance estimates are associated with major patterns of differentiation in the microbiota across people.**
32. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
- This study found a strong correlation between microbiota diversity and long-term diets as assessed using diet inventories.**
33. Loftus, E. V. Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterology* **126**, 1504–1517 (2004).
 34. Cann, H. M. *et al.* A human genome diversity cell line panel. *Science* **296**, 261–262 (2002).
 35. Bach, J. F. & Chatenoud, L. The hygiene hypothesis: an explanation for the increased frequency of insulin-dependent diabetes. *Cold Spring Harb. Perspect. Med.* **2**, a007799 (2012).
 36. Clayton, T. A., Baker, D., Lindon, J. C., Everett, J. R. & Nicholson, J. K. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl Acad. Sci. USA* **106**, 14728–14733 (2009).
 37. Jackson, R. L., Greiwe, J. S. & Schwen, R. J. Emerging evidence of the health benefits of S-equol, an estrogen receptor beta agonist. *Nutr. Rev.* **69**, 432–448 (2011).
 38. Setchell, K. D. & Clerici, C. Equol: history, chemistry, and formation. *J. Nutr.* **140**, 1355S–1362S (2010).
 39. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
 40. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
 41. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc. Natl Acad. Sci. USA* **108**, 4554–4561 (2011).
- This paper gives insight into the resilience of the human microbiota in the face of repeated disturbances, and the degree of baseline variation.**
42. Jakobsson, H. E. *et al.* Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS ONE* **5**, e9836 (2010).
 43. Beisner, B. E., Haydon, D. T. & Cuddington, K. Alternative stable states in ecology. *Front. Ecol. Environ.* **1**, 376–382 (2003).
 44. Walker, B., Hollin, C. S., Carpenter, S. R. & Kinzig, A. Resilience, adaptability and transformability in social-ecological systems. *Ecol. Soc.* **9**, <http://www.ecologyandsociety.org/vol9/iss2/art5/> (16 September, 2004).
 45. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
 46. Sun, Y. *et al.* Advanced computational algorithms for microbial community analysis using massive 16S rRNA sequence data. *Nucleic Acids Res.* **38**, e205 (2010).
 47. Knights, D., Parfrey, L. W., Zaneveld, J., Lozupone, C. & Knight, R. Human-associated microbial signatures: examining their predictive value. *Cell Host Microbe* **10**, 292–296 (2011).
 48. Carroll, I. M. *et al.* Molecular analysis of the luminal- and mucosal-associated intestinal microbiota in diarrhea-predominant irritable bowel syndrome. *Am. J. Physiol. Gastrointest. Liver Physiol.* **301**, G799–G807 (2011).
 49. Chang, J. Y. *et al.* Decreased diversity of the fecal microbiome in recurrent *Clostridium difficile*-associated diarrhea. *J. Infect. Dis.* **197**, 435–438 (2008).
 50. Young, V. B. & Schmidt, T. M. Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota. *J. Clin. Microbiol.* **42**, 1203–1206 (2004).
 51. Willing, B. P. *et al.* A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* **139**, 1844–1854 (2010).

52. Swidsinski, A., Loening-Baucke, V. & Herber, A. Mucosal flora in Crohn's disease and ulcerative colitis — an overview. *J. Physiol. Pharmacol.* **60**, 61–71 (2009).
53. Lozupone, C. *et al.* Identifying genomic and metabolic features that can underlie early successional and opportunistic lifestyles in human gut symbionts. *Genome Res.* <http://dx.doi.org/10.1101/gr.138198.112> (4 June, 2012).
54. Libby, J. M., Donta, S. T. & Wilkins, T. D. *Clostridium difficile* toxin A in infants. *J. Infect. Dis.* **148**, 606 (1983).
55. Yamamoto-Osaki, T., Kamiya, S., Sawamura, S., Kai, M. & Ozawa, A. Growth inhibition of *Clostridium difficile* by intestinal flora of infant faeces in continuous flow culture. *J. Med. Microbiol.* **40**, 179–187 (1994).
56. Folke, C. *et al.* Regime shifts, resilience, and biodiversity in ecosystem management. *Annu. Rev. Ecol. Evol. Syst.* **35**, 557–581 (2004).
57. Scheffer, M. *et al.* Floating plant dominance as a stable state. *Proc. Natl Acad. Sci. USA* **100**, 4040–4045 (2003).
58. Hazen, T. C. *et al.* Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science* **330**, 204–208 (2010).
59. Valentine, D. L. *et al.* Dynamic autoinoculation and the microbial ecology of a deep water hydrocarbon irruption. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1108820109> (10 January, 2012).
60. Jernberg, C., Lofmark, S., Edlund, C. & Jansson, J. K. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J.* **1**, 56–66 (2007).
61. van der Waaij, D., Berghuis, J. M. & Lekkerkerk, J. E. Colonization resistance of the digestive tract of mice during systemic antibiotic treatment. *J. Hyg. (Lond.)* **70**, 605–610 (1972).
62. McNulty, N. P. *et al.* The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. *Sci. Transl. Med.* **3**, 106ra106 (2011).
63. Manichanh, C. *et al.* Reshaping the gut microbiome with bacterial transplantation and antibiotic intake. *Genome Res.* **20**, 1411–1419 (2010). **This study indicates that the indigenous microbiota may be more plastic than previously thought. The observation that antibiotic pretreatment interfered with, rather than promoted, establishment of the donor community indicates that low species abundance or diversity alone cannot predict low colonization resistance.**
64. Levine, J. M. & D'antonio, C. M. Elton revisited: a review of evidence linking diversity and invasibility. *Oikos* **87**, 15–26 (1999).
65. Khoruts, A., Dicksved, J., Jansson, J. K. & Sadowsky, M. J. Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent *Clostridium difficile*-associated diarrhea. *J. Clin. Gastroenterol.* **44**, 354–360 (2010).
66. Gough, E., Shaikh, H. & Manges, A. R. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection. *Clin. Infect. Dis.* **53**, 994–1002 (2011).
67. Hautier, Y., Niklaus, P. A. & Hector, A. Competition for light causes plant biodiversity loss after eutrophication. *Science* **324**, 636–638 (2009).
68. Elmqvist, T. *et al.* Response diversity, ecosystem change, and resilience. *Front. Ecol. Environ.* **1**, 488–494 (2003).
69. Hansen, E. E. *et al.* Pan-genome of the dominant human gut-associated archaeon, *Methanobrevibacter smithii*, studied in twins. *Proc. Natl Acad. Sci. USA* **108**, 4599–4606 (2011).
70. Flint, H. J., Duncan, S. H., Scott, K. P. & Louis, P. Interactions and competition within the microbial community of the human colon: links between diet and health. *Environ. Microbiol.* **9**, 1101–1111 (2007).
71. Louis, P. *et al.* Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J. Bacteriol.* **186**, 2099–2106 (2004).
72. Chaffron, S., Rehrauer, H., Pernthaler, J. & von Mering, C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* **20**, 947–959 (2010).
73. Stecher, B. *et al.* Like will to like: abundances of closely related species can predict susceptibility to intestinal colonization by pathogenic and commensal bacteria. *PLoS Pathogens* **6**, e1000711 (2010).
74. Bever, J. D., Westover, K. M. & Antonovics, J. Incorporating the soil community into plant population dynamics: the utility of the feedback approach. *J. Ecol.* **85**, 561–573 (1997).
75. Stark, P. L. & Lee, A. The microbial ecology of the large bowel of breast-fed and formula-fed infants during the 1st year of life. *J. Med. Microbiol.* **15**, 189–203 (1982).
76. Glover, L. E. & Colgan, S. P. Hypoxia and metabolic factors that influence inflammatory bowel disease pathogenesis. *Gastroenterology* **140**, 1748–1755 (2011).
77. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **108**, 6252–6257 (2011).
78. Dupont, H. L. Gastrointestinal infections and the development of irritable bowel syndrome. *Curr. Opin. Infect. Dis.* **24**, 503–508 (2011).

Acknowledgements We would like to thank L. Parfrey, J. Knight and A. Knight for their comments on this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/716f2m. Correspondence should be addressed to R.K. (rob.knight@colorado.edu).

Reciprocal interactions of the intestinal microbiota and immune system

Craig L. Maynard¹, Charles O. Elson², Robin D. Hatton¹ & Casey T. Weaver¹

The emergence of the adaptive immune system in vertebrates set the stage for evolution of an advanced symbiotic relationship with the intestinal microbiota. The defining features of specificity and memory that characterize adaptive immunity have afforded vertebrates the mechanisms for efficiently tailoring immune responses to diverse types of microbes, whether to promote mutualism or host defence. These same attributes can put the host at risk of immune-mediated diseases that are increasingly linked to the intestinal microbiota. Understanding how the adaptive immune system copes with the remarkable number and diversity of microbes that colonize the digestive tract, and how the system integrates with more primitive innate immune mechanisms to maintain immune homeostasis, holds considerable promise for new approaches to modulate immune networks to treat and prevent disease.

Every one of us enters the world devoid of microbial colonization because of the sterile environment of the womb. This germ-free existence is short-lived: birth exposes the newborn to the microbiota of the mother, setting in motion the colonization of mucosal tissues in the digestive, respiratory and urogenital tracts, and the skin by a diverse microbiota, which we coexist with throughout our lives. The complex and dynamic interaction between the microbiota and its human host is the culmination of nearly half a billion years of co-evolution with vertebrates that has reciprocally shaped the repertoires of the microbiota and the immune system, such that the microbiota in humans is normally restrained and well-tolerated. The scope of this interaction is particularly evident in the intestinal tract, in which the greatest diversity and abundance of microbes reside. Estimated at approximately 100 trillion organisms, most of which are bacteria (although archaea and eukaryotes are also represented), the microbiota numbers about ten times the total cells in the human body, with the greatest density populating the distal ileum and colon¹. The collective genome, or metagenome, of the intestinal microbiota has more than 100 times the number of genes of the human genome. Each individual is populated by roughly 15% of the 1,000 or more species of intestinal bacteria that have been described², which reflects the substantial variability in the composition of the microbiota between individuals. There are, therefore, about tenfold more genes in each of our microbiomes than in each of us, encoding the greatest source of potential antigens for the immune system to cope with, substantially exceeding those of self and pathogen-derived antigens.

The relationship that has been forged between the intestinal microbiota and its human host provides mutual benefits. At homeostasis, the microbiota benefits from the warm, nutrient-rich environment of the gut so it can establish a relatively stable ecosystem. Humans in turn benefit from a highly adaptive metabolic engine that in addition to providing essential non-nutrient factors, such as vitamins, also substantially increases our ability to harvest nutrients from food. This increased digestive capacity is mainly a result of the microbiota's complementation of the limited diversity of complex-carbohydrate-metabolizing enzymes that are encoded in the human genome. In addition, by establishing robust, interlinked metabolic or nutrient networks, and biofilms among its constituents, the microbiota limits the resources available to potential pathogens that must outcompete well-adapted and entrenched resident microbes for metabolic

and physical niche space. Resident microbes thereby establish a microbial buffer that limits access by those not part of the consortium.

However, the microbiota is not innocuous, and under conditions that compromise the ability of the host to limit the microbiota's entry from the intestinal lumen, some species can invade host tissues to cause disease. Furthermore, shifts in the composition of the microbiota, whether induced by dietary changes, antibiotic treatment or invasive pathogens, can disturb the balance of organisms in the microbiota and alter the metabolic network of the collective to favour the outgrowth of potentially pathogenic constituents. Referred to as dysbiosis, such changes in the microbiota can perturb immune regulatory networks that normally restrain intestinal inflammation, and may contribute to immune-mediated disease directed against antigens of the microbiota. Dysbiosis is most often associated with inflammatory bowel disease (IBD), including Crohn's disease and ulcerative colitis, and necrotizing enterocolitis in premature infants. But it is also increasingly linked to a number of extraintestinal immune-mediated diseases, including rheumatoid arthritis, multiple sclerosis, diabetes, atopic dermatitis and asthma, as well as obesity and metabolic syndrome, all of which could have their pathogenic origins in untoward reactivity of the immune system to the microbiota. Whether dysbiosis is a cause or effect of these disorders remains to be determined, but understanding the factors that lead to alterations in the composition of the microbiota in these conditions promises to be informative, irrespective of their basis.

In this Review, we highlight advances in our understanding of the immune mechanisms by which the dynamic interplay of the intestinal microbiota and its host normally favours a homeostatic, mutualistic relationship, as a basis for understanding the causes of breakdowns in this relationship that lead to disease.

Co-evolution of the microbiota and adaptive immunity

Vertebrate evolution coincided with the emergence of the adaptive immune system, the main components of which are T and B lymphocytes. The acquisition of genetic recombinatorial mechanisms for generating diverse, anticipatory antigen-recognition receptors on T cells and B cells allowed specific responses to a vast diversity of antigens, and long-lived immune memory. Although evolutionary pressures that have shaped immune strategies are often viewed in the context of host defence, it has been proposed that the emergence of adaptive immunity

¹Department of Pathology, The University of Alabama at Birmingham, Birmingham, Alabama, USA; ²Department of Medicine, The University of Alabama at Birmingham, Birmingham, Alabama, USA.

could have been driven as a means to foster, rather than limit, microbial colonization³. In this view, a substantial survival advantage would have been afforded to organisms that could harness the extended and more flexible metabolic capacity derived from a permanent, diverse intestinal microbiota, as long as the attendant infectious risk could be mitigated. Adaptive immunity might have allowed this.

By adding new layers — such as secretory IgA (sIgA) — to existing innate barrier defences, early vertebrates with adaptive immune mechanisms could begin to promote colonization of their alimentary tracts by exerting selective pressures that favoured resident microbes that were relatively innocuous, but metabolically useful. The survival benefits that accrued as a result of derivation of nutrients from a broader range of foods could have been a major factor in the evolutionary success of vertebrates³. The emergence of adaptive immunity provided a means to recognize, and remember, both beneficial and detrimental members of the intestinal microbiota to foster maintenance of beneficial species at the expense of detrimental ones. In doing so, the emerging adaptive immune system would have had to acquire mechanisms for tempering innate immune responses programmed only for the clearance of microbes. The development of a broad repertoire of immune cells that could suppress, as well as promote, innate inflammatory mechanisms contingent on the threat level of the microbe they recognized would become invaluable.

Support for this hypothesis comes from the study of the diversity of the microbiota in invertebrates, which seems to be far less complex than the microbiota of vertebrates³. Although data on comparative metagenomics are limited, the highest density of bacteria found in either organisms or the environment is in the human colon¹. The intestinal microbiota of vertebrate species as divergent as rodents and zebrafish share similar complexity to that of humans, and their gastrointestinal tracts are similarly colonized soon after birth⁴. Further, a homologue of sIgA was recently identified in bony fish⁵. Now extending to all jawed vertebrate taxa, except cartilaginous fish and reptiles, this unique class of immunoglobulin is adapted for interactions with the commensal microbiota and mucosal pathogens. It is tempting to speculate that the development of multimeric mucosal IgA, and adaptations that effect its efficient transport across the mucosal epithelium, and the compartmentalization of the gut-associated lymphoid tissues (GALT) from the peripheral immune system were essential adaptations that allowed vertebrates to harness a complex intestinal microbiota. Although more work is needed to establish a comparative map of the microbiota in the digestive tracts of more vertebrate and non-vertebrate species, it is anticipated that the merger of comparative studies in metagenomics and immunology will yield insight into the adaptive mechanisms that have allowed vertebrates to become an evolutionary success. It may yet prove that we owe it to the bugs that have become 'us'.

The microbiota and the developing immune system

The mammalian immune system is perhaps the most elaborate example of the complex symbiotic relationship that has resulted from the co-evolution of vertebrates and their microbiota. Unlike other vertebrates, placental mammals give birth to live young that are carried to term *in utero* and are nursed with milk that is rich in maternal antibodies. These antibodies provide passive transfer of immunity from mother to infant, which has implications for the developing immune system of the infant, and the microbiota that colonize the gut.

Maternal effects on the neonatal microbiome

The greatest initial contribution to the composite human–microbiota 'superorganism' is the vertical transmission of components of the mother's microbiota to the child at birth. Normally, intestinal colonization of neonates is dominated by transmission of bacteria from the maternal vaginal flora, which is less diverse than that of the lower intestinal tract⁶. The 'pioneer' species received from the mother seem to be important, because infants born by Caesarean section — who are initially colonized by bacterial species of epidermal, rather than vaginal, origin — are predisposed to development of allergies and asthma later in life⁷. This is

despite the fact that, although Caesarean-section-born infants lag behind those born transvaginally in their acquisition of the two bacterial divisions dominant in the adult microbiota (Firmicutes and Bacteroidetes), they do catch up. Conversely, mother-to-child transmission in neonates delivered transvaginally rapidly shifts to skin and oral ecologies postpartum, such that these infants are later exposed to similar species as those delivered through Caesarean section. Thus, first contact could be deterministic. That is, pioneer bacterial species might have substantial and lasting effects on the immune response, irrespective of the composition of the mature microbiota.

The neonatal microbiota varies erratically until about 1-year-old when it stabilizes, establishing a consortium that resembles that of adults⁸. During this initial period, the neonatal immune system rapidly matures under the influence of the microbiota. Although environmental factors such as diet, exposure to new microbes, xenobiotics, bacteriophages and intestinal infections have important roles in shaping the composition of the microbiota during this maturational window, the part the neonatal immune system plays is less clear. What is clear is the initial and ongoing influence of breastfeeding on the infant's microbiome. In addition to a unique mix of nutrients and antimicrobial proteins that influence the ecology of the neonatal microbiota, breast milk provides abundant sIgA, the specificities of which have been shaped by the maternal microbiota. The mucosal immune memory of the mother is thereby transmitted to her offspring. Thus, in breastfed infants delivered transvaginally, the intestinal microbiota is not only seeded by maternal bacterial species, but its composition may also be reinforced and shaped by the maternal sIgA repertoire that is influenced, in turn, by the maternal microbiota.

Maturation of the infant mucosal immune system takes months, so the passive transfer of maternal sIgA has a considerable protective role against potential pathogens that could perturb the ecological trajectory of the infant's intestinal microbiota. Maternal sIgA also shields the neonatal immune system from its own microbiota, perhaps so the neonate's defences are not overwhelmed before they are fully developed. In so far as microbial antigens that are bound by sIgA are handled by the innate immune system in a 'tolerogenic' mode, partly owing to IgA's poor fixation of complement, transfer of maternal sIgA to the infant seems to favour the establishment of regulatory immune networks in the infant that promote a mutualistic relationship with the microbiota. Thus, in addition to the direct effect of maternal sIgA on the microbiota of the infant, it also has immunomodulatory effects on the developing infant's immune repertoire so as to indirectly influence its microbiota. The maternal–neonate metagenomic bond is, therefore, extended postnatally, providing one mechanism by which microbial ecologies tend to cluster in family members.

Microbiome effects on developing host–barrier defence

Maturation of the intestinal mucosa and its GALT — Peyer's patches of the distal ileum, isolated lymphoid follicles (ILFs) and mesenteric lymph nodes (MLNs) — is initiated by, and contingent on, intestinal colonization. Peyer's patches and MLNs develop prenatally, but ILFs develop postnatally⁹. However, each of these lymphoid tissues requires signals derived from the sensing of intestinal microbiota for their complete development, recruitment of a mature complement of immune cells, or both. Similarly, non-lymphoid structures of the intestinal mucosa that contribute to the establishment of host–microbiota mutualism are driven by colonization of the neonate.

Preparation for the adaptive immune response to neonatal colonization requires the prenatal actions of a subset of innate lymphoid cells (ILCs), termed lymphoid tissue inducer (LTi) cells (Fig. 1). LTi cells are instrumental in prenatal organization of the development of lymphoid tissues, including the components of the GALT. LTi cells develop in the fetal liver from a common lymphoid precursor that gives rise to all lymphoid cells. During fetal development, LTi cells disseminate to the MLN and Peyer's patches anlagen, stimulating the development of these structures, and the recruitment and partitioning of B and T cells into B-cell follicles and T-cell zones that characterize secondary lymphoid tissues^{9,10}.

The development of ILFs is also dependent on LTi cells, but is initiated

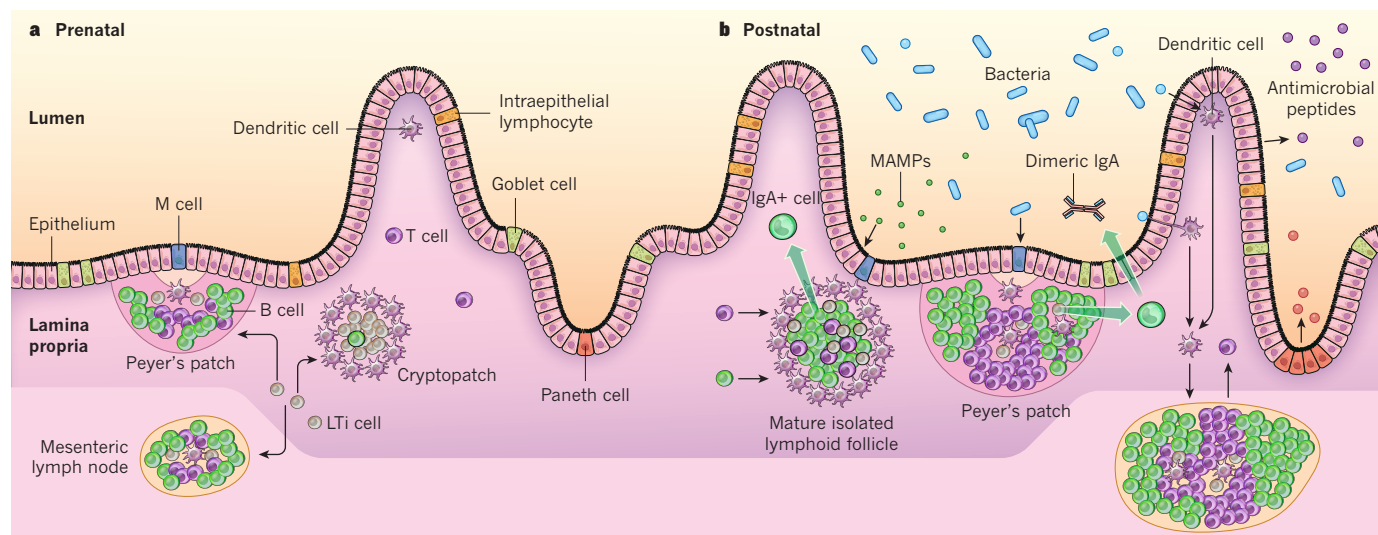


Figure 1 | The gut-associated lymphoid tissue establishes perinatal host-microbiota mutualism in the intestine. **a**, Prenatally, secondary lymphoid tissues (Peyer's patches and mesenteric lymph nodes) and cryptopatches develop by the spatiotemporal recruitment of lymphoid tissue inducer (LTi) cells to sites of the developing intestine and supporting neurovascular structures. This, in turn, stimulates the recruitment of dendritic cells, T cells and B cells in preparation for the immune response to the microbiota. Intraepithelial lymphocytes seed the epithelium before birth. **b**, Postnatally, bacteria colonize the neonatal intestine immediately, initiating multiple events that affect the development or functional maturation of the mucosa and gut-associated lymphoid tissues. Shown from left to right: microbe-associated molecular patterns (MAMPs) sensed by pattern-recognition receptors on intestinal epithelial cells and dendritic cells adjacent to cryptopatches stimulate the further recruitment of B cells and T cells, causing the cryptopatches to develop into mature isolated lymphoid follicles. The isolated lymphoid follicles release IgA-producing plasma cells — which are formed through

T-cell-dependent and independent interactions — into the lamina propria. Microbes also cross the epithelium and enter Peyer's patches through M cells, from which they are endocytosed by dendritic cells in the subepithelial dome. Antigen-loaded dendritic cells in the Peyer's patch interact with local lymphocytes to induce T-cell differentiation and T-cell-dependent B-cell maturation in the germinal centre to induce the development of IgA-producing plasma cells that home to the lamina propria, where they release dimeric IgA for transport into the intestinal lumen. Dendritic-cell-mediated luminal sampling of microbial products or transcytosis of bacteria across the epithelium results in antigen loading of lamina propria dendritic cells, which then migrate through the afferent lymphatics vessels (not shown) to a draining mesenteric lymph node to induce differentiation of effector T cells that traffic to the lamina propria. Shown on the far right, sensing of MAMPs stimulates the proliferation of intestinal epithelial cells in crypts, resulting in their increased depth and, in the small intestine, increased density of Paneth cells. This sensing also arms the intestinal epithelial cells for release of antimicrobial peptides.

after birth when these cells cluster in the lamina propria below the intestinal crypts, forming cryptopatches¹¹. ILFs develop from cryptopatches only after colonization by the intestinal microbiota. Central to sensing the colonizers of the intestinal tract is the expression of a diverse range of germline-encoded pattern-recognition receptors (PRRs) by intestinal epithelial cells (IECs) and immune cells resident in the gut. These receptors include transmembrane Toll-like receptors (TLRs) and C-type lectin receptors (CLRs) that reside on the cell surface or in endosomes, and cytosolic nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs), which cells of the intestinal mucosa use to detect microbe-associated molecular patterns (MAMPs) that are expressed by constituents of the resident microbiota, as well as pathogens. MAMPs are recognized — through mechanisms that are not well understood — by cells in the neonatal gut, stimulating ILF development to generate a lymphoid structure that is capable of supporting the maturation of B cells that produce sIgA. At least one of the microbial products that induces this process is peptidoglycan derived from Gram-negative bacteria. Recognition of peptidoglycan by NOD1 in IECs elicits production of CCL20 and β -defensin 3 that direct the recruitment of B cells to LTi-dendritic-cell clusters in cryptopatches to induce the expression of sIgA (ref.12).

Unlike MLNs and Peyer's patches, for which class-switch recombination to produce IgA is T-cell dependent, IgA class-switching in ILFs can occur through T-cell-dependent and -independent pathways¹³. In addition, MLNs and Peyer's patches are unique to mammals, whereas ILFs are found in non-mammalian vertebrates, suggesting that they are the evolutionary forerunners of the secondary lymphoid tissues of the GALT¹⁴ and may have evolved to accommodate, and promote, an increasingly complex intestinal microbiota before specialized secondary lymphoid tissues emerged. The importance of ILFs in regulating the microbiota is evident in mice that are devoid of these structures, and in

which Gram-negative bacteria are over-represented¹². Similar to Peyer's patches, ILFs develop in the intestinal mucosa in intimate association with the overlying intestinal epithelium (Fig. 1). Both structures are strategically deployed to sample the microbiota through specialized microfold, or M, cells, which transport intact microbes or their products across the epithelial barrier. In this way, the innate and adaptive immune systems constantly monitor the microbiota and are 'informed' of the dominant bacterial species that reside in proximity to the epithelium.

Like the GALT, the epithelium of the fetal and newborns' intestine is incompletely adapted to dense bacterial colonization, but responds rapidly to perinatal colonization to accommodate the microbiota. Expression of — and signalling by — TLR4, which is the receptor for the abundant product of Gram-negative bacteria, lipopolysaccharide, are significantly increased antecedent to intestinal colonization at birth¹⁵. Within hours of exposure to the microbiota, the response of IECs to lipopolysaccharide is markedly attenuated through the down-regulation of TLR4 and components of its signalling apparatus. This increase in TLR4 expression and signalling followed by down-regulation seems to represent a mechanism whereby the fetal epithelium is primed to transmit transient inflammatory signals on sensing the microbiota to promote an early immune response, but then becomes rapidly desensitized as an adaptation to the mounting bacterial load.

An additional immune mechanism that prepares the neonate for commensal colonization is the seeding of the intestinal epithelium with specialized intraepithelial lymphocytes before birth. Intraepithelial lymphocytes intercalate between IECs where they exist in an activated state and can respond rapidly to microbial encroachment. Throughout life, intestinal intraepithelial lymphocytes help to maintain the integrity of the epithelial-cell barrier, limit bacterial translocation and facilitate epithelium repair after injury by secreting soluble mediators, such as antimicrobial peptides¹⁶.

Although the host has evolved elaborate developmental strategies to prepare for postnatal colonization, the microbiota reciprocates by inducing maturation of the host's immune system. Therefore, compared with conventionally raised mice, germ-free or 'germ-reduced' mice have reduced size and cellularity of secondary lymphoid tissues as well as altered numbers, frequencies or diversity of immune-cell populations at mucosal sites, and thus mount abnormal responses to infection and injury. In addition, a postnatal time window is thought to exist in which regulatory T (T_{reg}) cells¹⁷ and invariant-natural killer T (iNKT) cells¹⁸ that have received their 'primary education' in the thymus, receive a 'secondary education' from the microbiota to establish a durable immune repertoire that is crucial to the prevention of inflammatory diseases in childhood and adulthood.

Innate pathways to sense and restrain the microbiota

The intestinal tract is the largest barrier tissue in the human body (it has a surface area of about 300 m² in adults), making it the most extensive portal for entry of commensal or pathogenic microbes. Highly specialized barrier defences have evolved to confine the microbiota and resist pathogens, while maintaining its main function of nutrient uptake. The strategy is one of a layered defence that integrates a stratified mucous layer, a relatively impenetrable but highly responsive epithelium, and a lamina propria populated by innate and adaptive immune cells that actively participate in homeostatic responses to restrain the microbiota without undue inflammation, yet are poised for the induction of antimicrobial clearance responses and tissue repair, should the barrier be breached.

Reciprocity between the epithelium and the microbiota

The epithelium is central to the orchestration of intestinal defences. Not simply a passive barrier to microbial translocation, the intestinal epithelium is an active sensor of, and conduit for, the dialogue between the host and microbiota. The five main cell types that comprise the epithelium — absorptive enterocytes, goblet cells, Paneth cells, M cells and enteroendocrine cells — develop from a common stem cell located near the base of the intestinal crypts. Each cell type has a specialized, integral role in intestinal homeostasis, and is both responsive to the microbiota and conditioned by it (Fig. 1).

Maintaining the polarized structure of the epithelial barrier is crucial to its function. The tight junctions that seal the interfaces of adjacent IECs segregate the epithelium into an apical, lumen-exposed surface and a basolateral surface anchored to the basement membrane. Depending on whether the membrane-associated PRRs are arrayed on the apical or basolateral surface of the IECs and on which side of the epithelial barrier MAMPs are detected, cells of the epithelium initiate responses that either promote the release of protective factors that are directed luminally (for example, secreted mucins and antimicrobial peptides) to directly restrain the microbiota, or internally (for example, cytokines

and chemokines) to either promote immune quiescence at homeostasis or activate inflammatory immune responses when the epithelial barrier has been breached.

As a mucosal tissue, the intestinal epithelium continuously produces and is invested by a layer of mucus that is a first line of defence against microbes (Fig. 2). Mucus is produced by goblet cells and is composed of heavily glycosylated mucin proteins, as well as other protective molecules, such as trefoil factor, that contribute to epithelial restitution and repair. Production of intestinal mucus is regulated by products of the microbiota. In germ-free mice the mucous layer in the colon is highly attenuated, despite normal numbers of mucin-laden goblet cells. Addition of the MAMP lipopolysaccharide or peptidoglycan stimulates the release of mucin by goblet cells and the rapid reconstitution of the colonic inner mucous layer¹⁹. Butyrate produced by benign constituents of the microbiota also promotes increased release of mucin, providing a positive-feedback loop for maintenance of the mucous barrier and its colonization by butyrate-producing commensals. The importance of the mucus layer is evident in mice deficient for principal intestinal mucin, Muc2. These mice have increased translocation of commensal and pathogenic bacteria²⁰, and spontaneously develop colitis²¹. Components of the healthy microbiota therefore directly contribute to the barrier function of the intestine through their induction of mucin production and secretion by goblet cells.

The thickness and continuity of intestinal mucus differs regionally: it is thinner and discontinuous in the proximal small intestine and becomes thicker and continuous in the distal small and large intestine²², showing some correlation with the local bacterial load (10^3 – 10^5 organisms per gram of luminal contents in the duodenum and jejunum; about 10^8 organisms per gram in the ileum, and 10^{10} – 10^{12} organisms per gram in the colon). The mucus is stratified into two functionally distinct layers: a compact, firmly adherent inner layer that is sparsely populated by bacteria and a more loosely structured, non-adherent outer layer that is at least tenfold more densely populated by the microbiota²².

Although both mucous layers have a similar Muc2-dominated composition, proteolytic cleavage of Muc2's polypeptide backbone in the outer layer results in its expanded volume and accessibility to colonization by components of the microbiota²². Indeed, the outer mucous layer provides an anchor for the attachment of bacteria of the microbiota that can establish biofilms that exclude pathogens. Furthermore, in addition to dietary glycans, mucin glycans are nutrients for some constituents of the microbiota, such as *Bifidobacterium* and *Bacteroides* spp., thereby promoting their retention in the collective (Box 1). These bacteria ferment complex O-linked mucin glycans to produce short-chain fatty acids (SCFAs) derived from mucin catabolism (for example, acetate and lactate) that are toxic to some pathogens²³, and produce other metabolites (such as, propionate and butyrate) that are the main nutrient source for colonic IECs. The SCFAs also signal through

BOX 1

Modulation of the microbiota by blood-group antigens

An important aspect of the microbiota's use of the intestinal mucous layer as an ecological niche is its decoration by blood-group antigens. Analogous to blood-group antigens on erythrocytes, the assembly of type A, B or Lewis-b glycans on intestinal mucins is contingent on the generation of the core H-glycan by the actions of FUT2. This protein is encoded by *FUT2*, which is functional in most individuals (secretor genotype), but is non-functional in a significant minority (about 20% of Caucasians, referred to as non-secretors) owing to a missense mutation. The presence or absence of a functional *FUT2* allele correlates strongly with the composition of the microbiota⁹⁷. In particular, *Bifidobacterium* spp., which are a beneficial component of the microbiota, are dependent on the terminal blood-group glycans

for colonization of the intestinal mucus and are less abundant in non-secretor genotypes. The non-secretor phenotype is associated with necrotizing enterocolitis and Gram-negative sepsis in premature infants⁹⁸, as well as those with Crohn's disease⁹⁹. Therefore, it seems that the protective benefits of colonization by *Bifidobacterium* spp. are a mutually beneficial evolutionary adaptation of this commensal to its host. Given the predisposition to dysbiosis and its deleterious health effects in non-secretors, it is likely that evolutionary pressure to retain the defective *FUT2* allele in the population is a result of mucin-linked blood-group glycans also serving as sites of attachment for mucosal viruses (such as noroviruses) so that those with a non-secretor genotype are protected¹⁰⁰.

G-protein-coupled receptors on IECs, to downregulate host inflammatory responses²⁴. Thus, colonization of the outer mucous layer by commensals is an important adaptation that supports a stable host–microbiota relationship.

In contrast with the outer mucous layer, the inner mucous layer provides a relatively impermeable barrier against the microbiota. This characteristic is a result of the layer's compact physicochemical structure and its function as a reservoir for microbicidal products of the epithelium, including antimicrobial peptides that are specialized to kill different classes of microbes and sIgA, which is retained in the mucous layer after being shuttled across the epithelium by polymeric immunoglobulin receptor (pIgR) (Fig. 2). Effectively, the inner mucous layer is a 'killing field' that few pathogens or commensals have evolved strategies to penetrate. This microbe-sparse zone is enforced, in part, by the antibacterial lectin Reg-IIIγ (ref. 25). Because Reg-IIIγ is selectively bactericidal for Gram-positive bacteria, it is likely that antimicrobial peptides specific for Gram-negative bacteria also contribute to the microbe-sparse zone. A similar spatial segregation of bacteria from the epithelium has been identified in the colon, in which a thicker inner mucous layer excludes most bacteria even in the face of a substantially higher bacterial load²².

The region of the intestinal tract with the poorest coverage of protective mucous, the proximal small intestine²², also has the greatest exposed epithelial surface area owing to its prominent villous structure. Although this feature favours digestion and absorption of nutrients it also seems to make the epithelium in this region particularly vulnerable to entry by the microbiota and pathogens. However, the bacterial loads in this region are the lowest along the length of the intestine, with over a million-fold fewer bacteria per unit of luminal contents than the large intestine. This is, in part, owing to the more vigorous peristaltic motility of the proximal small intestine, which rapidly clears material, including microbes, from the lumen. The lumen of the proximal small intestine is also the entry point for the contents of the gall bladder and stomach, which contain high levels of bile salts and acid that have antimicrobial effects.

In addition, the bases of the crypts of the small intestine are home to many Paneth cells. These IECs are arrayed with a range of PRRs and are specialized for the production and release of abundant antimicrobial peptides, including α -defensins, which are small, highly cationic microbicides unique to Paneth cells. In contrast with other antimicrobial peptides, such as Reg-IIIγ, synthesis of which requires signals from the microbiota²⁶, α -defensins are synthesized and stored in Paneth-cell granules without the need for sensing of MAMPs. However, the release of Paneth-cell granules is induced by MAMPs, and the secretion of active α -defensins has been shown to control the composition of the microbiota²⁷. *NOD2*, which was the first susceptibility gene linked to Crohn's disease^{28,29}, encodes an NLR that is important in sensing the microbiota to control the release of antimicrobial peptides by Paneth cells. Deficiencies of antimicrobial peptide production, such as occur in *NOD2* mutants, are thought to result in altered microbiota composition and density in the small intestine that heighten susceptibility to intestinal inflammation, particularly in the terminal ileum where the highest density of Paneth cells are found.

Colonic IECs also regulate the composition of microbiota through PRR-dependent mechanisms. Mice deficient in the NLRP6

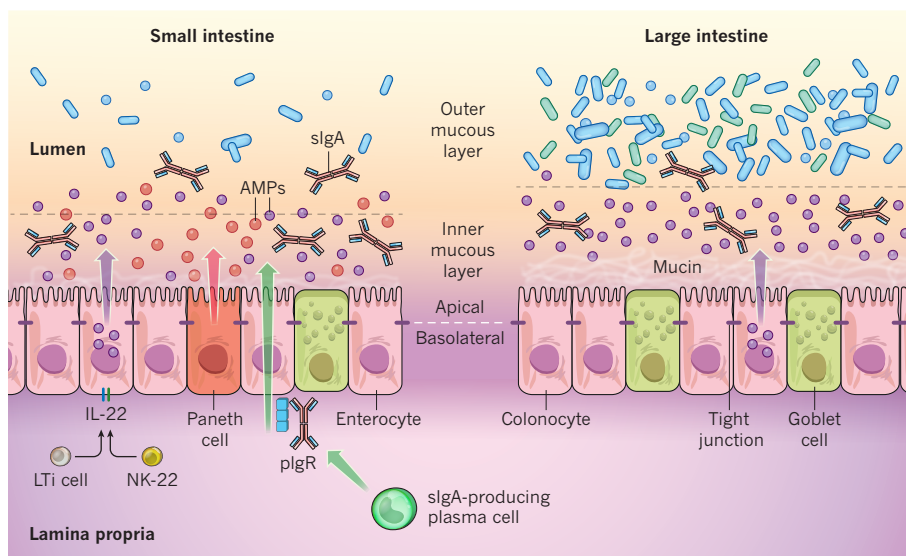


Figure 2 | The barrier function of the intestinal epithelium. Distinct subpopulations of intestinal epithelial cells (IECs) are integrated into a continuous, single cell layer that is divided into apical and basolateral regions by tight junctions. Enterocytes, in the small intestine, and colonocytes in the large intestine, as well as specialized Paneth cells in the bases of small intestinal crypts continually sense the microbiota to induce the production of antimicrobial peptides (AMPs). Goblet cells produce mucin, that is organized into a dense, more highly cross-linked inner proteoglycan gel that forms an IEC-adherent inner mucous layer, and a less densely cross-linked outer mucous layer. The outer layer is highly colonized by constituents of the microbiota. The inner mucous layer is largely impervious to bacterial colonization or penetration due to its high concentration of bactericidal AMPs, as well as commensal-specific secretory IgA (sIgA), which is ferried across IECs from their basolateral surface, where it is bound by the polymeric immunoglobulin receptor (pIgR), to the inner mucous layer where it is released by proteolytic cleavage of pIgR. Responding to the microbiota, innate lymphoid cells, including RORγt- and AhR-expressing LTi and NK-22 cells, produce IL-22, which stimulates AMP production and promotes epithelial barrier integrity.

inflammasome in colonic IECs have an altered microbiota that confers increased susceptibility to colitis because of damage to the colonic epithelium³⁰. The increased susceptibility to colitis that results from NLRP6 deficiency is transmissible to wild-type mice, indicating that a dysbiotic flora is a contributory factor. How deficiency of the NLRP6 inflammasome results in an altered microbiota is incompletely defined, although reduced interleukin (IL)-18 levels in NLRP6-deficient mice suggests this cytokine has an important role.

Reciprocity between the epithelium and innate immune cells

Although the intestinal mucous layer largely insulates the intestinal epithelium from direct interactions with the microbiota, bacterial metabolites and components are able to permeate this zone and alter gene expression in IECs through PRRs³¹. In addition to the epithelial products that are secreted apically to restrict contact with the microbiota (for example, mucins and antimicrobial peptides), the epithelium also produces factors, such as chemokines and cytokines, that are secreted basolaterally. These factors signal immune cells residing internal of the epithelium, particularly in the intestinal lamina propria (Fig. 3). Activation of PRRs typically promotes pro-inflammatory innate responses, so the intestinal epithelium had to evolve strategies to mitigate these responses for commensals such that, at homeostasis, cytokine signals transmitted to mucosal immune cells limit inflammation. Several mechanisms by which this anti-inflammatory state is favoured have emerged.

The strategic distribution of TLRs and NLRs on and within IECs has a considerable effect on whether bacterial MAMPs will be recognized, and if so, whether their recognition promotes pro-inflammatory responses or represses them. Pathogenic bacteria possess virulence factors to allow them to attach to or invade IECs thereby introducing MAMPs into the IEC cytosol where they are recognized by NLRs; however, bacterial strains of the indigenous microbiota are non-invasive and therefore less potent activators of NLRs. In the intact epithelium,

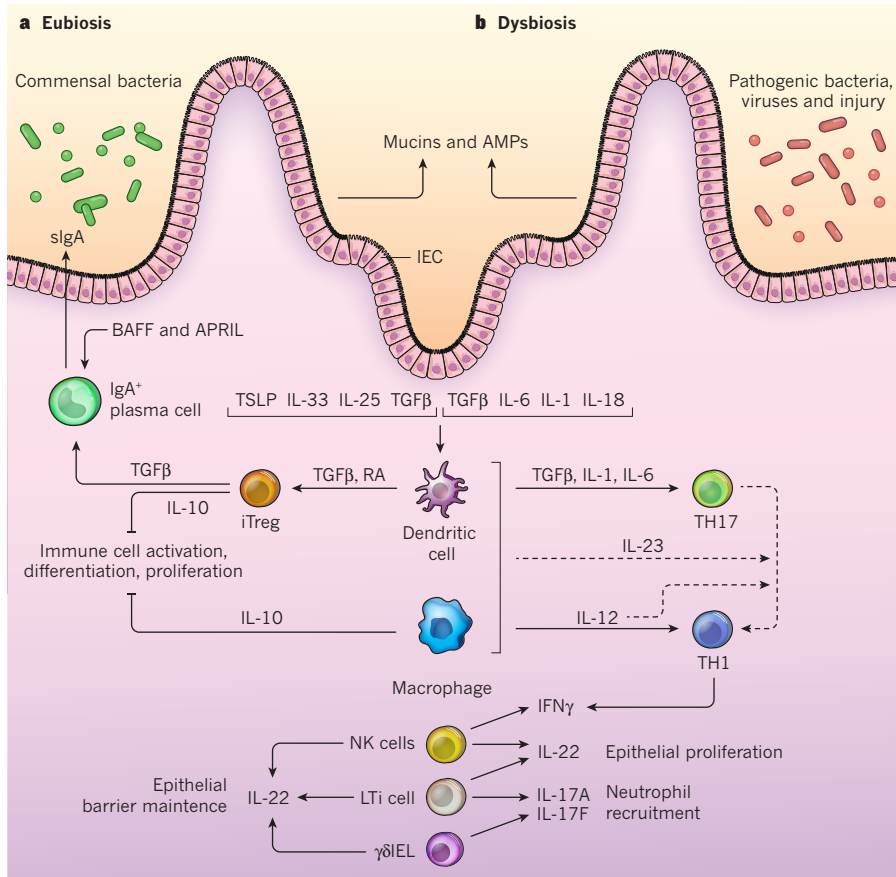


Figure 3 | The epithelial-innate-adaptive continuum in response to microbial antigens. **a.** In response to the microbiota, IECs secrete mucins and AMPs that limit microbial interaction with epithelial cells. Under homeostatic, eubiotic conditions, MAMPs stimulate the secretion of cytokines (including TSLP, IL-33, IL-25, and TGFβ) by IECs that promote development of tolerogenic macrophages and dendritic cells. Dendritic cells, in turn, induce the development of induced T_{reg} (iT_{reg}) cells through a TGFβ- and retinoic acid (RA)-dependent process. Through multiple mechanisms, including the secretion of TGFβ and IL-10 by iT_{reg} cells and the secretion of IL-10 by macrophages, the anti-inflammatory balance of the intestine is maintained by inhibiting or dampening potential effector responses. In addition, T_{reg} cell-derived TGFβ and epithelial-derived BAFF and APRIL promote development of IgA⁺ plasma cells to ensure an abundant supply of sIgA in the lumen that further limits microbial interaction with the epithelium. **b.** In the face of pathogen invasion, mucosal injury or dysbiosis, MAMPs stimulate the secretion of pro-inflammatory cytokines by IECs (including, IL-6, IL-1 and IL-18) and intestinal dendritic cells and macrophages (including IL-6, IL-23 and IL-12) that induce development of the effector CD4⁺ T cells T_H1 and T_H17, the latter of which can transition to the former as a result of IL-23 or IL-12 signalling. Intestinal innate lymphoid cells, including NK-like cells, LTi cells, and γδIELs, respond to pro-inflammatory cytokines to upregulate IL-22, which helps to protect the epithelial barrier, and IL-17A and IL-17F, which are involved in neutrophil recruitment.

MAMPs derived from commensals are largely restricted to interactions with apically accessible TLRs, which seem to be functionally dampened by tonic exposure to the microbiota, whether through decreased expression, impaired signalling or both.

By contrast, expression of TLRs on the basolateral aspects of IECs can signal a barrier breach by both commensals and pathogens. This is exemplified by TLR5, the receptor for flagellin, which is expressed on the basolateral aspects of the epithelium where it can detect the repeated flagellin monomers that make up bacterial flagellae³². Detection of flagellin on the 'host' side of the intestinal epithelial barrier results in epithelial-cell activation, which initiates inflammatory responses by innate and adaptive immune cells with the aim of eradicating the invading bacterium. The importance of flagellin as an immune-activating MAMP and an antigen targeted by adaptive immune cells is reflected in its high representation among antigens bound by serum antibodies from colitic mice and in patients with Crohn's disease³³.

Apical expression of TLRs and restricted access to cytosolic NLRs are adaptations to constrain interactions with luminal microbes, but some commensals have also evolved strategies that actively dampen TLR signalling in IECs, should contact occur. *Bacteroides* and *Lactobacillus* spp. inhibit activation of the classical NF-κB pathway, which is central to induction of pro-inflammatory gene expression downstream of all PRRs. Active NF-κB dimers are sequestered in the cytosol by association with IκB complexes. Release of NF-κB for nuclear localization is contingent on the phosphorylation of IκB, by which it is targeted for ubiquitylation and proteasomal degradation downstream of PRR signalling. Contact of IECs with these commensals inhibits IκB degradation, thereby blocking nuclear transport of NF-κB (ref. 34). *Bacteroides* spp. can further dampen NF-κB signalling by enhancing the nuclear export of NF-κB by inducing increased expression of the nuclear receptor family member, peroxisome-proliferation-activated receptor-γ (PPAR-γ). PPAR-γ binds NF-κB in the nucleus and shuttles it back to the cytoplasm, thereby terminating its transactivation of pro-inflammatory

genes³⁵. Thus, in addition to host mechanisms that normally restrain direct contact between commensals and IECs, co-adaptation of the microbiota and host have provided mechanisms that suppress pro-inflammatory engagement of PRRs in the intact epithelium.

The epithelium also supports tolerance of the microbiota by conditioning intestinal dendritic-cells through directional release of immunomodulatory cytokines into the lamina propria. At homeostasis, IECs secrete thymic stromal lymphopoietin (TSLP), IL-33 and IL-25, which promote tolerogenic activities of a subset of intestinal dendritic-cells defined by expression of integrin α_Eβ₇ complex (also known as CD103)³⁶. Antigens presented by CD103⁺ dendritic-cells favour the development of T_{reg} cells and IgA⁺ plasma cells that 'home' to the lamina propria at which they repress inflammatory responses to the microbiota³⁷. IECs also produce abundant TGF-β, which suppresses NF-κB-dependent pro-inflammatory signalling in intestinal macrophages and dendritic cells, and promotes development and maintenance of T_{reg} cells and IgA⁺ plasma cells. TLR-activated IECs can directly promote T-independent development of sIgA producing plasma cells in ILFs by the release of both B-cell activating factor (BAFF) and a proliferation-inducing ligand (APRIL)³⁸. In addition to cytokines, IECs express other factors that limit pro-inflammatory responses of mucosal immune cells, including products of tryptophan metabolism and prostaglandins (for example, PGE₂).

Like IECs, intestinal macrophages normally express lower levels of TLRs and are hyporesponsive to TLR signalling to dampen inflammatory responses to the commensal microbiota. This is despite their retention of highly active phagocytic and bactericidal activities³⁹. Also like IECs, this inflammatory anergy is probably multifactorial, but reflects the limited entry of commensal bacteria or their products into the cytosol in which they can activate NLRs — despite their uptake and degradation in phagolysosomes. Although phagocytosis of commensal bacteria induces pro-IL-1β in intestinal macrophages, its cleavage by caspase-1 to produce the active form requires NLRP3- or NLRP4-containing inflammasomes, which are activated by MAMPs

that are more effectively delivered into the cytosol by pathogenic, rather than commensal, bacteria⁴⁰. Therefore, phagocytes in the intestinal lamina propria are conditioned to mount inflammatory responses against pathogens that deliver MAMPs to cytosolic sensors, but to efficiently clear commensals in a non-inflammatory manner.

Innate lymphoid-cells response to epithelial barrier breach

Although LT α cells are indispensable in the development of the GALT, they are only one subset of a growing family of ILCs that have emerged as important contributors to intestinal homeostasis and barrier defence in the postnatal immune system^{41,42}. The diversity of ILC subsets seems to parallel that of CD4⁺ T cells, with which they share many functional features. However, the antigen-specific-receptor-lacking ILCs are recruited to mucosal immune responses mainly by cytokines, which are produced by IECs or intestinal myeloid cells (dendritic cells and macrophages).

ILCs are distributed in the intestinal lamina propria and GALT, in which they are poised to respond rapidly to microbes that penetrate the epithelial defences. LT α cells are particularly important to barrier defence against bacterial incursions⁴³, as are a subset of natural-killer-like cells (referred to as NK-22 or ILC22 cells), which seem to share many features with LT α cells⁴⁴. Common to these cells, and to CD4⁺ T cells of the T H 17 lineage (discussed later) is their expression of the transcription factor retinoic acid-related orphan receptor (ROR) γ t, which is required for these cells' development and function⁹, and production of the cytokines IL-22, IL-17A and IL-17F (ref. 45). IL-22 is a member of the IL-10 cytokine family that acts on epithelial cells of barrier tissues, including the intestine, to enhance antimicrobial defence and epithelial barrier integrity. IL-17A and IL-17F target innate immune cells, stromal cells and epithelial cells to induce cytokines and chemokines (for example, G-CSF and CXCL8) that induce increased neutrophil production and recruitment, respectively.

ROR γ t⁺ ILCs seem to be the main source of IL-22 at steady state and, in a continuous feedback loop with the microbiota, release IL-22, IL-17 or both in response to microbiota-induced production of IL-23 by macrophages and dendritic cells. These cytokines stimulate epithelial-cell secretion of antimicrobial peptides that inhibit or kill bacteria in the vicinity of the epithelial cell surface. Although ILCs can contribute to intestinal pathology under certain circumstances⁴⁶, they are crucial in the early host response to enteropathogenic bacteria^{41,47}.

ROR γ t⁺ ILCs also express the aryl hydrocarbon receptor (AhR) with which they sense microbiota metabolites and xenobiotics^{45,48}. AhR signalling is required both for ILC maintenance and IL-22 production⁴⁸. Akin to its essential role in GALT development, ILC production of lymphotoxin provides an important amplifying loop for the production of IL-22 by ILCs, stimulating dendritic cell production of IL-23 that, in turn, increases IL-22 production by ILCs^{49,50}. In addition to ROR γ t and AhR, another transcription factor shared between innate and adaptive immune cells that influences the composition of the microbiota is T-bet. This transcription factor is involved in T H 1 cell development, and is also expressed by a distinct subset of ILCs. Under pathogen-free conditions, RAG-deficient mice, which lack B and T cells, are relatively resistant to inflammation driven by their microbiota, despite the absence of an adaptive immune system. This is largely owing to compensatory increases in ILCs. However, when deficient for T-bet, RAG-deficient mice develop spontaneous intestinal inflammation resembling ulcerative colitis (known as T-bet^{-/-} RAG^{-/-} ulcerative colitis (TRUC) mice)⁵¹. Disease is a result of overexpression of TNF α by dendritic cells and an absence of T reg cells. Remarkably, transfer of the microbiota from TRUC mice to immunocompetent mice transfers disease, reflecting the outgrowth of two pathogenic bacterial strains: *Klebsiella pneumoniae* and *Proteus mirabilis*. Although details of the mechanisms responsible for T-bet-dependent restraint of the microbiota remain to be defined, the dysbiosis in TRUC mice highlights the importance of innate immune cells in regulating the composition of the microbiota and is evidence that a dysbiotic flora can drive disease in otherwise healthy hosts.

Adaptive immunity in homeostatic responses to microbiota

The interplay between the microbiota, intestinal epithelium and innate and adaptive immune cells at homeostasis favours the dominance of regulatory networks that prevent inflammation or immune-mediated disease. Remarkably, the innate defences are sufficiently robust that much of the host response to the microbiota progresses without involvement of CD4⁺ T-cell-dependent responses. Furthermore, a portion of the sIgA response that contributes to partitioning of commensal organisms away from the epithelium is generated in ILFs without the requirement for T cells¹³. When the adaptive immune response is recruited, it is typically limited to the mucosal tissues such that systemic immunity is not generated. This compartmentalization of mucosal immunity is a result of the programming of adaptive immune cells to travel back to the mucosae following their differentiation in Peyer's patches or MLNs. Finally, at steady state, antigen presentation by intestinal dendritic cells favours the development of regulatory CD4⁺ T cells, which suppress the development of pro-inflammatory innate and effector T-cell responses to avert excessive inflammation.

Regulatory T-cell responses to the microbiota

CD4⁺ T reg cells are essential for maintenance of mutualism with the microbiota. The frequencies of these T cells are considerably elevated in the intestine relative to other tissues⁵², and microbiota-induced T reg cells affect the composition of the microbiota⁵³. The crucial role of T reg cells in immune homeostasis to the microbiota is well-documented by the consequences of these cells' absence⁵⁴. T reg -cell deficiency, irrespective of the means by which it is induced, results in unopposed effector T-cell responses and IBD, driven by reactivity to antigens of the microbiota. Indeed, much of our understanding of the inflammatory potential of the commensal microbiota has been garnered through models in which regulatory pathways have been disrupted⁵⁴. The largely non-overlapping TCR specificities of regulatory and effector T-cell subsets in the colonic lamina propria and associated lymphoid tissues suggest that distinct antigenic epitopes, whether derived from the same or distinct members of the microbiota, control these competing T-cell fates⁵⁵.

Regulatory CD4⁺ T cells include subsets that are distinguished on the basis of their expression of the transcription factor, Foxp3. Foxp3⁺ (T reg) or Foxp3⁻ (T regulatory cells 1 (Tr1)) are each characterized by their production of IL-10, one of the main immunoregulatory cytokines required for immune tolerance of the intestinal microbiota. The non-redundant role of IL-10 in intestinal immune homeostasis is well-established; IL-10-deficient mice develop spontaneous, unremitting colonic inflammation driven by IL-23 and the T H 17 pathway^{56,57}. The essential role of TLR in sensing the microbiota in this process is evident by the absence of disease in germ-free IL-10-deficient mice and mice deficient for both IL-10 and MyD88 (ref. 58, 59). Although non-T cells produce IL-10, T-cell-derived IL-10 is crucial for intestinal immune homeostasis. Thus, IL-10 deficiency that is restricted to total CD4⁺ T cells results in spontaneous colitis that has a severity comparable with that in mice with global loss of IL-10 (ref. 60). Although deficiency of IL-10 that is limited to Foxp3⁺ T reg cells also induces colitis, its severity is reduced, indicating that Foxp3⁺ CD4⁺ T cells contribute to the protective IL-10 response.

Myeloid cells of the innate immune system are the main targets for regulatory actions of IL-10 (ref. 61), although IL-10 signalling in CD4⁺ T cells also seems to be contributory^{62,63}. Importantly, polymorphisms in the IL-10 locus confer susceptibility to IBD in humans⁶⁴, and early-onset IBD has been linked to mutations in IL-10 receptor components that impair signalling⁶⁵.

Remarkably, IL-10-producing T cells can be induced to develop in response to specific commensals or their products. The capsular polysaccharide-A moiety of the common commensal *Bacteroides fragilis* mediates interaction of *B. fragilis* with the colonic mucosa, facilitating both colonization and activation of an anti-inflammatory cascade⁶⁶. Polysaccharide-A promotes expansion of colonic IL-10-expressing Foxp3⁺ T reg cells, through the TLR2–MyD88 pathway⁶⁶. The

expression of TLR2 on CD4 T-cells suggests there is the capacity for direct regulation by microbial products that favours polysaccharide-A-dependent T_{reg} -cell responses over T_H1 responses⁶⁷, although there are other known TLR2 ligands that do not promote induction of Foxp3 or IL-10 (ref. 66).

A cocktail of *Clostridium* species, mostly comprised of clusters IV and XIVa, drives the efficient expansion of T_{reg} cells in the colons of germ-free mice, seemingly owing to their superior ability to elicit TGF β production in the intestinal mucosa⁶⁸. In contrast with induction of T_{reg} cells by *B. fragilis*, *Clostridium*-dependent T_{reg} expansion occurs independently of MyD88 through mechanisms yet to be defined⁶⁸. To date, there are no studies using otherwise complete microbiota that are devoid of *B. fragilis* or this specific collection of *Clostridium* spp., leaving open the issue of the level of redundancy in this process. However, *B. fragilis* is absent from the microbiota of 20–30% of humans, indicating the existence of other pathways to T_{reg} -cell development. Accordingly, default induction or expansion of T_{reg} cells can be achieved by colonization of germ-free mice with a defined altered Schaedler flora (ASF) that consists of eight species of the microbiota dominated by *Bacteroides distasonis*, but devoid of *B. fragilis*⁶⁹. Thus, although there seems to be a proclivity of certain components of the microbiota to promote T_{reg} -cell development, this does not seem to be a unique property of any one microbial component, reflecting substantial redundancy in the system to ensure that T_{reg} -cell dominance is established.

Effector T-cell responses to the microbiota

Similar to T_{reg} cells and B cells, the largest deployment of effector CD4⁺ T-cells in the healthy immune system is in the intestines, and is largely driven by the microbiota. Although the intestines mount robust T_H2 responses to helminth infestations, in countries where these infections are less common the main effector T-cell subsets resident in the gut express cytokines characteristic of T_H17 and T_H1 cells.

T_H17 cells in particular, which are thought to be the most ancient of effector T-cell subsets⁴, seem to have evolved to bolster mucosal barrier defences to promote mutualism with the microbiota. T_H17 cells share developmental requirements for ROR γ t and AhR (ref. 70) with intestinal ILCs, as well as a similar effector cytokine repertoire, including IL-17A, IL-17F and IL-22. Furthermore, T_H17 cells have developmental ties with T_{reg} cells through their mutual requirement for the abundant intestinal cytokine TGF β (ref. 4). Finally, the T_H17 developmental pathway is distinguished by considerable plasticity⁷¹, allowing divergent functional programs contingent on local pro- or anti-inflammatory cues. Although a last line of defence, effector CD4⁺ T-cells are perpetually engaged in the host–microbiota dialogue by favouring regulatory benefits at homeostasis or pathological consequences when dysregulated.

The contribution of the microbiota to intestinal T_H17 -cell development is highlighted by the virtual absence of this subset in germ-free mice⁷². In addition to IL-6, which is required for T_H17 differentiation and its deviation away from induced T_{reg} programming, multiple microbiota-dependent factors favour T_H17 development in the intestines, including TGF β , IL-1 β (ref. 73), IL-23 (ref. 74) and even ATP derived from commensal bacteria⁷⁵. Similar to the ability of a limited quorum of commensal bacterial species to disproportionately influence induced T_{reg} -cell development, minor constituents of the microbiota can amplify intestinal T_H17 cell numbers. In mice, an unusually potent, but not unique, inducer of T_H17 cells⁶⁹ is the *Clostridium* sp. *Candidatus arthromitus* or segmented filamentous bacteria (SFB)^{72,76}. This bacterium populates the ileum and caecum and has long been known to be a potent activator of intestinal immune responses⁷⁷. Induction of T_H17 cells by SFB provides protection against gut pathogens⁷⁸, suggesting that general amplification of T_H17 effector cells can be host-protective. However, T_H17 induction by SFB is not entirely benign, because mono-association of mice with SFB induces T_H17 -mediated inflammatory arthritis⁷⁹ and multiple-sclerosis-like symptoms in the experimental autoimmune encephalomyelitis (EAE) model⁸⁰. Remarkably, therefore, sensing of a single constituent of the intestinal microbiota can promote

autoimmunity in extraintestinal tissues. At present, it is unclear whether SFB, or related organisms, exert similar effects in humans.

Whether at steady state, responding to infection or chronically inflamed as a result of IBD, the intestine contains, in addition to effectors that express IL-17, T cells that express IFN γ or both IL-17 and IFN γ . Of note, T_H17 cells retain a high capacity for divergent cytokine expression profiles and function, or plasticity, after their commitment to the T_H17 pathway. Specifically, T_H17 cells can give rise to IFN γ producers that resemble classical T_H1 cells. This has been demonstrated in humans and mice^{71,81}, and has suggested that many of the T_H1 cells found in the intestine arise from the T_H17 pathway (Fig. 3). The relative contributions of T_H1 cells to immune protection or pathogenesis remain to be elucidated, whether they are derived from the T_H17 pathway or not.

Balancing regulatory and effector responses

To balance accommodation of the microbiota and the need to mount host defences against microbial invasion, the adaptive immune system has evolved to direct the development of distinct CD4⁺ T-cell fates by different APC subsets and the microbe-induced factors they produce. Although T_{reg} -cell dominance is the default at homeostasis, the shared dependence of induced T_{reg} and T_H17 -cell development on TGF β provides an elegant means to alternately divert programming of naive CD4⁺ T cells from a homeostatic, microbe-tolerant response to an inflammatory, microbe-clearing response contingent on the cofactors integrated with TGF β signalling⁴ (Fig. 3).

At homeostasis, naive CD4⁺ T-cell recognition of antigens derived from the microbiota favours induced T_{reg} -cell development by virtue of the production, by intestinal dendritic cells, of the vitamin A metabolite all-trans-retinoic acid^{182–84}. Vitamin A is not synthesized by the host so its influence on the regulatory tone of the mucosal immune response is predicated from adequate dietary intake. Retinoic acid is derived from vitamin A by the sequential actions of the ubiquitous enzyme, alcohol dehydrogenase and one of three retinal-specific aldehyde dehydrogenases, which are more restricted in their tissue distribution, including CD103⁺ dendritic cells and IECs in the intestine. The potency of retinoic acid as a cofactor for TGF β -dependent development of induced T_{reg} cells — and suppression of T_H17 development — and its constitutive production by a regulatory subset of intestinal dendritic cells at steady state reflect the robust default pathway to T-cell-mediated regulation. Under conditions in which microbial antigens promote production of pro-inflammatory cytokines by dendritic cells, retinoic acid is either repressed⁸⁵ or co-opted by pro-inflammatory pathways^{86,87} to override T_{reg} -cell induction.

In common with T_H17 cells, and perhaps reflecting an aspect of shared TGF β -dependent programming that favours multipotency, induced T_{reg} cells display developmental plasticity that influences adaptive immunity to the microbiota. When acted on by pro-inflammatory cytokines, induced T_{reg} cells can down-regulate Foxp3 and up-regulate ROR γ t (ref. 88), and T_{reg} cells can be converted into IL-17- or IFN γ -expressing effectors during infection⁸⁹. Notably, although Foxp3⁺ T_{reg} cells may be diverted to effector responses, the reverse is less apparent; established effector T cells, including T_H17 cells, seem to be more resistant to reprogramming to become Foxp3⁺ regulatory cells. Accordingly, microbial antigens that are identified as products of a pathogen are likely to be remembered as such through their imprinting of an effector T-cell response. By contrast, microbial antigens that promote induced T_{reg} responses at homeostasis, may, in the context of an inflammatory response, reprogram induced T_{reg} cells to generate an effector T-cell response that breaks the ‘tolerance’ to that microbe.

In a further example of the developmental plasticity of T_{reg} cells, recent studies report their role in IgA class-switch recombination in the gut, representing another arm of T_{reg} cell function in promoting immune homeostasis to the microbiota. TGF β has long been recognized as the principal switch factor for development of IgA-producing B cells. In a new twist on TGF β 's role in this process, induced T_{reg} cells were found to down-modulate Foxp3 and acquire features of follicular helper T cells

that promoted IgA production in Peyer's patches⁹⁰. Depletion of T_{reg} cells also caused a rapid loss of IgA⁺ plasma cells and sIgA production in the intestine⁹¹, indicating that T_{reg} cells participate both in the development and maintenance of intestinal IgA⁺ B cells. Although known to be important in host defence against infections, the main role of sIgA is to establish mutualism with the intestinal microbiota, as evidenced by its marked depletion in germ-free mice and systemic IgG response to the microbiota in animals that specifically lack sIgA. Coupled with the fact that about 75% of sIgA reactive to the microbiota develops through the T-cell-dependent pathway^{91,92}, this suggests that sIgA is mainly regulated by intestinal T_{reg} cells. In addition to providing evidence of further plasticity in the T_{reg} developmental program, these findings identify an additional link between adaptive immune networks that cope with the commensal flora, and extend the long-established role of TGFβ in promoting host-microbiota mutualism⁴.

Dysregulated immunity to the microbiota

Owing to their defined antigenic specificity and durable memory, effector CD4⁺ T cells are a liability when inappropriately directed against self-antigens or, in the case of the microbiota, the 'extended' self. In view of the enormous number of antigens expressed in the intestinal metagenome, it is remarkable that dysregulated effector responses to the microbiota are the exception. When they do occur, the result is IBD, including Crohn's disease or ulcerative colitis, which have distinct clinical and histopathological features, but a common requirement for the intestinal microbiota. Similar to autoimmunity directed against host self-antigens, dysregulated CD4⁺ T-cell responses to antigens of the microbiota leads to chronic, typically relapsing and remitting disease that reflects the immune system's inability to eliminate the antigens that drive the abnormal response.

That being said, inflammatory immune responses in the intestines result in detectable alterations of the resident microbiota, irrespective of the cause⁹³. This presents a challenge in our attempts to understand how certain microbes positively or negatively affect the disease process in IBD because it is difficult to draw conclusions without knowing the composition of the microbiota before diagnosis. As such, our understanding of the antigenic components of the microbiota that are the targets of immune attack in IBD are limited. However, experimental mouse models and patients with Crohn's disease show a remarkably shared reactivity to specific flagellin epitopes expressed by *Clostridium* spp., which seem to be immunodominant despite their minor representation within the commensal flora^{33,94}. What attributes of these commensals make them particularly common targets in a large subset of patients with Crohn's disease — but not ulcerative colitis — are unknown. However, it is likely to reflect unique properties that place them in specific geographical niches in the microbiota or in intimate contact with the immune system where their expression of flagella, which is less typical for commensal bacteria, make them unusually proficient at inducing effector rather than regulatory responses. Irrespective of the basis, retention of these organisms in the collective despite the pathogenic risk they pose implies there is a benefit to host-microbiota mutualism that has been evolutionarily conserved and remains to be understood.

In view of the prominent role of the T_H17 pathway in adaptive immunity to the intestinal microbiota and propensity for generating inflammation promoted by both IL-17 and IFNγ, it is not surprising that this pathway has emerged as a major contributor to IBD pathogenesis. Prior to the discovery of T_H17, Crohn's disease and ulcerative colitis were viewed in the context of T_H1- and T_H2-centric mechanisms, respectively; but, increasingly, data from genome-wide association studies implicate the contributions of the T_H17 pathway in both diseases⁹⁵. The discovery of IL-23 revolutionized views on the immunopathogenesis of autoimmunity and led to discovery of T_H17 cells⁹⁶. It is therefore fitting that this was the first cytokine linked to IBD by genome-wide association studies; variants of its receptor (IL-23R) have been found to confer both protection and susceptibility to Crohn's disease and ulcerative colitis⁹⁵.

As the number of genome-wide association and next-generation

sequencing studies has proliferated, so has the number of genes linked to the T_H17 pathway in IBD (ref. 95). This has been supplemented by genes of innate and adaptive immune pathways that are integrated with the T_H17 response and control it, spanning the gamut from epithelial barrier integrity maintenance and restitution to microbial sensing, to immunomodulatory cytokines. Because individual susceptibility alleles typically confer minor risk, multiple genetic susceptibility alleles are generally required for disease. Efforts to define this clustering of risk alleles is increasingly defining genotypes that confer substantial risk, as well as identifying interlaced susceptibility pathways that identify immunopathogenetic subsets of IBD — including those that do not directly implicate the T_H17 pathway. However, as more IBD susceptibility genotypes are found to overlap with extraintestinal immune-mediated diseases that do share a T_H17 pathogenesis, they increasingly implicate a role for the microbiota in conferring disease risk beyond the confines of the gut. As we advance into the era of integrated metagenomics of the human-microbiota 'superorganism', the opportunities for personalized medicine that were originally envisioned from sequencing of the human genome may well multiply in proportion to the increase in genes contributed by our microbiota, with attendant opportunities for more specifically targeted treatments and prevention of diseases that have their origins in the merger of vertebrates and their microbiomes millions of years ago. ■

- Ley, R. E., Peterson, D. A. & Gordon, J. I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837–848 (2006).
- Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
- McFall-Ngai, M. Adaptive immunity: care for the community. *Nature* **445**, 153 (2007).
- Weaver, C. T. & Hatton, R. D. Interplay between the T_H17 and T_{reg} cell lineages: a (co-)evolutionary perspective. *Nature Rev. Immunol.* **9**, 883–889 (2009).
- Zhang, Y.-A. *et al.* IgT, a primitive immunoglobulin class specialized in mucosal immunity. *Nature Immunol.* **11**, 827–835 (2010).
- Dominguez-Bello, M. G. *et al.* Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc. Natl Acad. Sci. USA* **107**, 11971–11975 (2010).
- Bager, P., Wohlfahrt, J. & Westergaard, T. Caesarean delivery and risk of atopy and allergic disease: meta-analyses. *Clin. Exp. Allergy* **38**, 634–642 (2008).
- Palmer, C., Bik, E., Digulio, D., Relman, D. & Brown, P. Development of the human infant intestinal microbiota. *PLoS Biol.* **5**, e177 (2007).
- Cherrier, M. & Eberl, G. The development of LT α i cells. *Curr. Opin. Immunol.* **24**, 178–183 (2012).
- van de Pavert, S. A. & Mebius, R. E. New insights into the development of lymphoid tissues. *Nature Rev. Immunol.* **10**, 664–674 (2010).
- Kanamori, Y. *et al.* Identification of novel lymphoid tissues in murine intestinal mucosa where clusters of c-kit⁺IL-7R⁺Thy1⁺ lympho-hemopoietic progenitors develop. *J. Exp. Med.* **184**, 1449–1459 (1996).
- Bouskra, D. *et al.* Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* **456**, 507–510 (2008).
- Tsuiji, M. *et al.* Requirement for lymphoid tissue-inducer cells in isolated follicle formation and T cell-independent immunoglobulin A generation in the gut. *Immunity* **29**, 261–271 (2008).
- Lane, P. J. L. *et al.* Lymphoid tissue inducer cells: bridges between the ancient innate and the modern adaptive immune systems. *Mucosal Immunol.* **2**, 472–477 (2009).
- Lotz, M. *et al.* Postnatal acquisition of endotoxin tolerance in intestinal epithelial cells. *J. Exp. Med.* **203**, 973–984 (2006).
- Ismaïl, A. S. *et al.* γδ intraepithelial lymphocytes are essential mediators of host-microbial homeostasis at the intestinal mucosal surface. *Proc. Natl Acad. Sci. USA* **108**, 8743–8748 (2011).
- Hansen, C. H. *et al.* Patterns of early gut colonization shape future immune responses of the host. *PLoS ONE* **7**, e34043 (2012).
- Olszak, T. *et al.* Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336**, 489–493 (2012).
- Petersson, J. *et al.* Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **300**, G327–G333 (2011).
- Bergstrom, K. S. B. *et al.* Muc2 protects against lethal infectious colitis by dissociating pathogenic and commensal bacteria from the colonic mucosa. *PLoS Pathogens* **6**, e1000902 (2010).
- Van der Sluis, M. *et al.* Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**, 117–129 (2006).
- Johansson, M. E. V., Larsson, J. M. H. & Hansson, G. C. The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc. Natl Acad. Sci. USA* **108**, 4659–4665 (2011).
- Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543–547 (2011).

This article reports comparative genomic and metabolic profiling of protective compared with non-protective strains of *Bifidobacterium* to identify a class of bacterial carbohydrate transporters that impart host-protective effects through the generation of acetate.

24. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009).
25. Vaishnava, S. *et al.* The antibacterial lectin Reg-III γ promotes the spatial segregation of microbiota and host in the intestine. *Science* **334**, 255–258 (2011).
26. Cash, H. L., Whitham, C. V., Behrendt, C. L. & Hooper, L. V. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126–1130 (2006).

This study demonstrated that Reg-III γ , a C-type lectin produced by Paneth cells in response to the resident microbiota, has microbicidal activity, identifying a new mechanism for sequestration of constituents of the microbiota to the intestinal lumen.

27. Salzman, N. H. *et al.* Enteric defensins are essential regulators of intestinal microbial ecology. *Nature Immunol.* **11**, 76–83 (2010).
28. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
29. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).

References 28 and 29 identified the first major susceptibility gene for IBD. These studies also implicated a link between impaired innate immune recognition of the microbiota and dysregulated adaptive immunity in IBD pathogenesis.

30. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757 (2011).
31. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by Toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).

This report established that TLR signalling in response to the resident microbiota has an important role in promoting maintenance of epithelial barrier function.

32. Vijay-Kumar, M., Aitken, J. D. & Gewirtz, A. T. Toll like receptor-5: protecting the gut from enteric microbes. *Semin. Immunopathol.* **30**, 11–21 (2008).
33. Lodes, M. J. *et al.* Bacterial flagellin is a dominant antigen in Crohn disease. *J. Clin. Invest.* **113**, 1296–1306 (2004).

This study reports an expression library that was made from the microbiota of spontaneously colitic mice and was screened with antibodies in their serum to identify commensal bacterial flagellins as major antigenic targets of the immune response in IBD.

34. Neish, A. S. *et al.* Prokaryotic regulation of epithelial responses by inhibition of I κ B- α ubiquitination. *Science* **289**, 1560–1563 (2000).
35. Kelly, D. *et al.* Commensal anaerobic gut bacteria attenuate inflammation by regulating nuclear–cytoplasmic shuttling of PPAR- γ and RelA. *Nature Immunol.* **5**, 104–112 (2004).
36. Hill, D. A. & Artis, D. Intestinal bacteria and the regulation of immune cell homeostasis. *Annu. Rev. Immunol.* **28**, 623–667 (2010).
37. Rescigno, M. & Di Sabatino, A. Dendritic cells in intestinal homeostasis and disease. *J. Clin. Invest.* **119**, 2441–2450 (2009).
38. He, B. *et al.* Intestinal bacteria trigger T cell-independent immunoglobulin A2 class switching by inducing epithelial-cell secretion of the cytokine APRIL. *Immunity* **26**, 812–826 (2007).
39. Smythies, L. E. *et al.* Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J. Clin. Invest.* **115**, 66–75 (2005).
40. Franchi, L. *et al.* NLR4-driven production of IL-1 β discriminates between pathogenic and commensal bacteria and promotes host intestinal defense. *Nature Immunol.* **13**, 449–456 (2012).
41. Colonna, M. Interleukin-22-producing natural killer cells and lymphoid tissue inducer-like cells in mucosal immunity. *Immunity* **31**, 15–23 (2009).
42. Spits, H. & Di Santo, J. P. The expanding family of innate lymphoid cells: regulators and effectors of immunity and tissue remodeling. *Nature Immunol.* **12**, 21–27 (2011).
43. Sonnenberg, G. F., Monticelli, L. A., Elloso, M. M., Fouser, L. A. & Artis, D. CD4 $^{+}$ lymphoid tissue-inducer cells promote innate immunity in the gut. *Immunity* **34**, 122–134 (2011).
44. Cella, M. *et al.* A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. *Nature* **457**, 722–725 (2009).
45. Takatori, H. *et al.* Lymphoid tissue inducer-like cells are an innate source of IL-17 and IL-22. *J. Exp. Med.* **206**, 35–41 (2009).
46. Buonocore, S. *et al.* Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* **464**, 1371–1375 (2010).
47. Sonnenberg, G. F., Fouser, L. A. & Artis, D. Border patrol: regulation of immunity, inflammation and tissue homeostasis at barrier surfaces by IL-22. *Nature Immunol.* **12**, 383–390 (2011).
48. Qiu, J. *et al.* The aryl hydrocarbon receptor regulates gut immunity through modulation of innate lymphoid cells. *Immunity* **36**, 92–104 (2012).
49. Ota, N. *et al.* IL-22 bridges the lymphotoxin pathway with the maintenance of colonic lymphoid structures during infection with *Citrobacter rodentium*. *Nature Immunol.* **12**, 941–948 (2011).
50. Tumanov, A. V. *et al.* Lymphotoxin controls the IL-22 protection pathway in gut innate lymphoid cells during mucosal pathogen challenge. *Cell Host Microbe* **10**, 44–53 (2011).
51. Garrett, W. S. *et al.* Enterobacteriaceae act in concert with the gut microbiota to induce spontaneous and maternally transmitted colitis. *Cell Host Microbe* **8**, 292–300 (2010).

52. Maynard, C. L. *et al.* Regulatory T cells expressing interleukin 10 develop from Foxp3 $^{+}$ and Foxp3 $^{-}$ precursor cells in the absence of interleukin 10. *Nature Immunol.* **8**, 931–941 (2007).
 53. Josefowicz, S. Z. *et al.* Extrathymically generated regulatory T cells control mucosal T $_{H}2$ inflammation. *Nature* **482**, 395–399 (2012).
 54. Izcue, A., Coombes, J. L. & Powrie, F. Regulatory lymphocytes and intestinal inflammation. *Annu. Rev. Immunol.* **27**, 313–338 (2009).
 55. Lathrop, S. K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478**, 250–254 (2011).
 56. Kühn, R., Löhler, J., Rennick, D., Rajewsky, K. & Müller, W. Interleukin-10-deficient mice develop chronic enterocolitis. *Cell* **75**, 263–274 (1993).
- This study identified a non-redundant function for IL-10 in immune homeostasis in the intestine, and observed the importance of composition of the microbiota in modulating disease severity in genetically susceptible mice.**
57. Yen, D. *et al.* IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *J. Clin. Invest.* **116**, 1310–1316 (2006).
 58. Sellon, R. K. *et al.* Resident enteric bacteria are necessary for development of spontaneous colitis and immune system activation in interleukin-10-deficient mice. *Infect. Immun.* **66**, 5224–5231 (1998).
 59. Rakoff-Nahoum, S., Hao, L. & Medzhitov, R. Role of Toll-like receptors in spontaneous commensal-dependent colitis. *Immunity* **25**, 319–329 (2006).
 60. Roers, A. *et al.* T cell-specific inactivation of the interleukin 10 gene in mice results in enhanced T cell responses but normal innate responses to lipopolysaccharide or skin irritation. *J. Exp. Med.* **200**, 1289–1297 (2004).
 61. Takeda, K. *et al.* Enhanced Th1 activity and development of chronic enterocolitis in mice devoid of Stat3 in macrophages and neutrophils. *Immunity* **10**, 39–49 (1999).
 62. Chaudhry, A. *et al.* Interleukin-10 signaling in regulatory T cells is required for suppression of Th17 cell-mediated inflammation. *Immunity* **34**, 566–578 (2011).
 63. Huber, S. *et al.* Th17 cells express interleukin-10 receptor and are controlled by Foxp3 $^{+}$ and Foxp3 $^{-}$ regulatory CD4 $^{+}$ T cells in an interleukin-10-dependent manner. *Immunity* **34**, 554–565 (2011).
 64. Franke, A. *et al.* Sequence variants in *IL10*, *ARPC2* and multiple other loci contribute to ulcerative colitis susceptibility. *Nature Genet.* **40**, 1319–1323 (2008).
 65. Glocker, E.-O. *et al.* Inflammatory bowel disease and mutations affecting the interleukin-10 receptor. *N. Engl. J. Med.* **361**, 2033–2045 (2009).
 66. Round, J. L. *et al.* The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* **332**, 974–977 (2011).
 67. Wang, Q. *et al.* A bacterial carbohydrate links innate and adaptive responses through Toll-like receptor 2. *J. Exp. Med.* **203**, 2853–2863 (2006).
 68. Atarashi, K. *et al.* Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* **331**, 337–341 (2011).
 69. Geuking, M. B. *et al.* Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity* **34**, 794–806 (2011).
 70. Sawa, S. *et al.* ROR γ^{+} innate lymphoid cells regulate intestinal homeostasis by integrating negative signals from the symbiotic microbiota. *Nature Immunol.* **12**, 320–326 (2011).
 71. Lee, Y. K. *et al.* Late developmental plasticity in the T helper 17 lineage. *Immunity* **30**, 92–107 (2009).
 72. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* **4**, 337–349 (2008).
 73. Shaw, M. H., Kamada, N., Kim, Y.-G. & Núñez, G. Microbiota-induced IL-1 β , but not IL-6, is critical for the development of steady-state TH17 cells in the intestine. *J. Exp. Med.* **209**, 251–258 (2012).
 74. Zaph, C. *et al.* Commensal-dependent expression of IL-25 regulates the IL-23–IL-17 axis in the intestine. *J. Exp. Med.* **205**, 2191–2198 (2008).
 75. Atarashi, K. *et al.* ATP drives lamina propria TH17 cell differentiation. *Nature* **455**, 808–812 (2008).
 76. Gaboriau-Routhiau, V. *et al.* The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009).
 77. Talham, G. L., Jiang, H. Q., Bos, N. A. & Cebra, J. J. Segmented filamentous bacteria are potent stimuli of a physiologically normal state of the murine gut mucosal immune system. *Infect. Immun.* **67**, 1992–2000 (1999).
 78. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).
 79. Wu, H.-J. *et al.* Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* **32**, 815–827 (2010).
- This study reported that a single component of the microbiota that induces increased intestinal T $_{H}17$ cells could drive the development of extraintestinal immune-mediated disease in genetically susceptible mice. This raises the possibility that antigen-independent effects of the mucosal immune response to limited constituents of the microbiota may trigger systemic autoimmunity.**
80. Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **108**, 4615–4622 (2011).
 81. Annunziato, F. *et al.* Phenotypic and functional features of human Th17 cells. *J. Exp. Med.* **204**, 1849–1861 (2007).
 82. Coombes, J. L. *et al.* A functionally specialized population of mucosal CD103 $^{+}$ DCs induces Foxp3 $^{+}$ regulatory T cells via a TGF- β and retinoic acid-dependent mechanism. *J. Exp. Med.* **204**, 1757–1764 (2007).
 83. Sun, C. M. *et al.* Small intestine lamina propria dendritic cells promote de novo generation of Foxp3 T reg cells via retinoic acid. *J. Exp. Med.* **204**, 1775–1785 (2007).

84. Mucida, D. *et al.* Reciprocal TH17 and regulatory T cell differentiation mediated by retinoic acid. *Science* **317**, 256–260 (2007).
 85. Maynard, C. L. *et al.* Contrasting roles for all-trans retinoic acid in TGF- β -mediated induction of *Foxp3* and *Il10* genes in developing regulatory T cells. *J. Exp. Med.* **206**, 343–357 (2009).
 86. Hall, J. *et al.* Commensal DNA limits regulatory T cell conversion and is a natural adjuvant of intestinal immune responses. *Immunity* **29**, 637–649 (2008).
 87. DePaolo, R. W. *et al.* Co-adjuvant effects of retinoic acid and IL-15 induce inflammatory immunity to dietary antigens. *Nature* **471**, 220–224 (2011).
 88. Lee, Y. K., Mukasa, R., Hatton, R. D. & Weaver, C. T. Developmental plasticity of Th17 and T_{reg} cells. *Curr. Opin. Immunol.* **21**, 274–280 (2009).
 89. Wohlfert, E. & Belkaid, Y. Plasticity of T_{reg} at infected sites. *Mucosal Immunol.* **3**, 213–215 (2010).
 90. Tsuji, M. *et al.* Preferential generation of follicular B helper T cells from Foxp3⁺ T cells in gut Peyer's patches. *Science* **323**, 1488–1492 (2009).
 91. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T. R. & Elson, C. O. A dominant, coordinated T regulatory cell–IgA response to the intestinal microbiota. *Proc. Natl Acad. Sci. USA* **106**, 19256–19261 (2009).
- References 90 and 91 identify a central role for T_{reg} cells in the induction and maintenance of IgA-producing plasma cells in the intestines, extending the functionality of T_{reg} cells in supporting mutualism to the microbiota.**
92. Macpherson, A. J. & Uhr, T. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science* **303**, 1662–1665 (2004).
 93. Lupp, C. *et al.* Host-mediated inflammation disrupts the intestinal microbiota and promotes the overgrowth of Enterobacteriaceae. *Cell Host Microbe* **2**, 119–129 (2007).
 94. Targan, S. R. *et al.* Antibodies to CBir1 flagellin define a unique response that is associated independently with complicated Crohn's disease. *Gastroenterology* **128**, 2020–2028 (2005).
 95. Khor, B., Gardet, A. & Xavier, R. J. Genetics and pathogenesis of inflammatory bowel disease. *Nature* **474**, 307–317 (2011).
 96. Kastelein, R. A., Hunter, C. A. & Cua, D. J. Discovery and biology of IL-23 and IL-27: related but functionally distinct regulators of inflammation. *Annu. Rev. Immunol.* **25**, 221–242 (2007).
 97. Wacklin, P. *et al.* Secretor genotype (*FUT2* gene) is strongly associated with the composition of bifidobacteria in the human intestine. *PLoS ONE* **6**, e20113 (2011).
 98. Morrow, A. L. *et al.* Fucosyltransferase 2 non-secretor and low secretor status predicts severe outcomes in premature infants. *J. Pediatr.* **158**, 745–751 (2011).
 99. McGovern, D. P. B. *et al.* Fucosyltransferase 2 (*FUT2*) non-secretor status is associated with Crohn's disease. *Hum. Mol. Genet.* **19**, 3468–3476 (2010).
 100. Rydell, G. E., Kindberg, E., Larson, G. & Svensson, L. Susceptibility to winter vomiting disease: a sweet matter. *Rev. Med. Virol.* **21**, 370–382 (2011).

Acknowledgements The authors thank D. Randolph and C. Morrow for discussions and critical review of this manuscript. C.T.W., C.O.E. and R.D.H. are supported by grants from the National Institutes of Health, and C.T.W. and C.L.M. are supported by grants from the Crohn's and Colitis Foundation of America. The authors extend their apologies to colleagues whose work could not be adequately acknowledged owing to space limitations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/nqbfui. Correspondence should be addressed to C.T.W. (cweaver@uab.edu).

Functional interactions between the gut microbiota and host metabolism

Valentina Tremaroli^{1,2} & Fredrik Bäckhed^{1,2,3}

The link between the microbes in the human gut and the development of obesity, cardiovascular disease and metabolic syndromes, such as type 2 diabetes, is becoming clearer. However, because of the complexity of the microbial community, the functional connections are less well understood. Studies in both mice and humans are helping to show what effect the gut microbiota has on host metabolism by improving energy yield from food and modulating dietary or the host-derived compounds that alter host metabolic pathways. Through increased knowledge of the mechanisms involved in the interactions between the microbiota and its host, we will be in a better position to develop treatments for metabolic disease.

Changes to lifestyle and an increase in the availability of energy-rich foods are important contributors to the world-wide obesity epidemic. The microbial inhabitants of the gut can also have an influence on metabolic processes, such as energy extraction from food, and should be considered an environmental factor that contributes to obesity and its comorbidities (such as insulin resistance, diabetes and cardiovascular disease).

Culture-independent methods to study microbial communities have advanced our knowledge of this human gut microbiota (Box 1) (see page 250 of this issue). Profiling of the common proxy for this community, the faecal microbiota, by 16S ribosomal RNA surveys and by direct sequencing of genetic material have shown that the human gut microbiota is a complex community of 100 trillion archaeal and bacterial cells distributed over more than 1,000 species¹ (Box 2). The community is dominated by bacteria, with more than 90% of the species belonging to Firmicutes and Bacteroidetes. Each person has a distinct and highly variable microbiota, but a conserved set of gut colonizers (the core gut microbiota) and genes (the core microbiome) are shared among individuals^{1,2} and may be required for the correct functioning of the gut.

Germ-free mice are those born and reared without exposure to any live microbes, and they provide a powerful tool for understanding the effects of the gut microbiota on host physiology. These mice can be colonized either with selected microbial species or whole communities from mice or humans to examine the transmissibility of physiological and pathological phenotypes, and to test what role the microbiota has in a particular phenotype. The gut microbiota in these mice modulates bone-mass density³ and promotes fat storage⁴, intestinal angiogenesis^{5,6} and the development of an immune response^{7,8} (see page 231 of this issue). In this Review, we discuss the metagenomic and gnotobiotic-based evidence for the role of the gut microbiota in energy metabolism and the possible links with obesity.

Obesity

Gut microbiota composition is altered in people who are obese, and it can respond to changes in body weight. Genetically obese *ob/ob* mice⁹ are hyperphagic as a result of a mutation in the gene that encodes the satiety-promoting hormone leptin. The caecal microbiota of these mice contains more Firmicutes and fewer Bacteroidetes than that of their lean wild-type littermates, even when

the mice are fed the same low-fat, polysaccharide-rich diet⁹. Similar changes have also been seen in the faecal microbiota of humans who are obese¹⁰. Bacteroidetes levels increase when weight is reduced, either by fat- or carbohydrate-restricted diets¹⁰, suggesting that Bacteroidetes may be responsive to calorie intake. A similar effect has also been observed in people who lost weight after a Roux-en-Y gastric bypass procedure. In these patients, increased levels of *Bacteroides* and *Prevotella* were negatively correlated with energy intake and adiposity¹¹. Other studies showed no such shift in the Firmicutes–Bacteroidetes ratio^{12–14}, but this may be because they used different clinical criteria (such as level of obesity, age, degree of weight loss and duration of calorie restriction), geographical locations, population sizes and microbiota-profiling methodologies. Although obesity and energy intake can affect the microbial composition, whether the gut microbiota contributes to obesity in humans is unclear.

A gastric bypass promotes sustained weight reduction and diminishes the risk of diabetes and cardiovascular disease for people who are obese^{15,16}. This knowledge has allowed the relationship between microbiota and obesity to be explored further. After a gastric bypass, diabetes can resolve before patients begin to lose weight, suggesting that this type of surgery has a direct antidiabetic effect. Exactly how this happens is not clear, but a shift in the composition of the faecal microbiota of humans^{11,14} suggests the gut microbiota contributes to the improved metabolic phenotype after a gastric bypass. The beneficial microbe *Faecalibacterium prausnitzii*, in particular, is less abundant in patients who are obese and diabetic, but increases after surgery¹¹. Levels of *F. prausnitzii* are negatively correlated with inflammatory markers, indicating that the bacterium may modulate systemic inflammation (common to diabetes and obesity) and contribute to the amelioration of diabetes. In addition, germ-free mice do not develop diet-induced obesity, and treatment of obese mice with antibiotics reduces adiposity and adipose inflammation, and improves glucose metabolism^{17–19}, further supporting the benefits of inducing a shift in microbiota composition.

Energy harvest

Carbohydrates are important sources of energy for human and microbial cells. Human enzymes cannot degrade most complex carbohydrates and plant polysaccharides. Instead, the non-digestible

¹Wallenberg Laboratory for Cardiovascular and Metabolic Research, Sahlgrenska University Hospital, 413 45 Gothenburg, Sweden. ²Department of Molecular and Clinical Medicine, University of Gothenburg, 413 45 Gothenburg, Sweden. ³Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health Sciences, University of Copenhagen, Copenhagen DK-2200, Denmark.

carbohydrates, including cellulose, xylans, resistant starch and inulin, are fermented in the colon by its microbiota to yield energy for microbial growth and end products such as short-chain fatty acids (SCFAs) (Fig. 1), mainly acetate, propionate and butyrate, which have profound effects on gut health as, for example, an energy source, an inflammation modulator, a vasodilator and part of gut motility and wound healing. In addition, SCFAs are energy substrates for the colonic epithelium (butyrate) and peripheral tissues (acetate and propionate)²⁰. The patterns of intestinal fermentation, and consequently the types and amount of SCFAs produced, are determined by how much carbohydrate is consumed and the composition of the gut microbiota. For example, fermentation of dietary fructans increases when gnotobiotic mice that have been colonized with *Bacteroides thetaiotaomicron*, are co-colonized with *Methanobrevibacter smithii*²¹. *B. thetaiotaomicron* produces more acetate and formate, and *M. smithii* uses formate for methanogenesis. The interactions promote more efficient carbohydrate fermentation and increased energy absorption from the gut, resulting in increased adiposity in the co-colonized mice compared with mice colonized with only *B. thetaiotaomicron*. The composition of the gut microbiota and the metabolic interactions between its species may therefore affect food digestion and energy harvest.

Direct evidence for the role of the microbiota in energy harvest and fat deposition comes from germ-free rats, which have reduced intestinal levels of SCFAs²², and twice as much urinary and faecal excretion of calories as that of conventional rats fed the same polysaccharide-rich diet²³. The germ-free rodents compensate for the reduced energy harvest by increasing their food intake²³. Germ-free mice also have reduced adiposity compared with their conventional counterparts, but adiposity is normalized when they are colonized with a healthy microbiota for 14 days^{4,19}. Microbial energy harvest in obesity has been investigated in conventional genetically obese *ob/ob* mice, which have increased amounts of SCFAs in their caecum and reduced energy content in their faeces compared with their lean littermates²⁴. Metagenomic sequencing of the caecal microbiota showed an enrichment of gene functions that were related to the degradation of dietary polysaccharides in the microbiome of *ob/ob* mice²⁴. This finding was also true of humans: the faecal microbiota of people who are obese has an increased capacity to harvest energy². In mice, the obese phenotype was transmissible through microbiota transplants, and germ-free mice colonized with the microbiota from obese donors gained twice as much fat as those colonized with the microbiota from lean donors²⁴.

The role of the gut microbiota in promoting energy harvest from diet and fat deposition has been clearly demonstrated in mice, but most of the evidence in humans has come from indirect studies. For instance, people who are obese have higher levels of ethanol in their breath than lean people²⁵, indicating altered fermentation and a greater number of faecal SCFAs¹³, which may suggest increased microbial energy harvest.

Diet alters the gut microbiota

Diet is known to modulate the composition of the gut microbiota in humans and mice. Long-term dietary habits have a considerable effect on the human gut microbiota. For example, children in a rural African village, who consumed high amounts of plant polysaccharides, had low levels of Firmicutes and increased levels of Bacteroidetes — mainly *Prevotella* and *Xylanibacter* — in their faecal microbiota compared with Italian children, who had high levels of Enterobacteriaceae — mainly *Shigella* and *Escherichia*²⁶. *Prevotella* and *Xylanibacter* are known to degrade cellulose and xylans, and are associated with increased faecal SCFAs, suggesting that the gut microbiota of the children living in rural Africa had adapted to maximize energy extraction from a diet rich in fibre. Human gut microbiota can be divided into three discrete compositions. However, this concept is currently being challenged as enterotypes may be more of

BOX 1

Terminology

● **Enterotype** is the grouping of the microbiota of a given person into discrete configurations. But recent data have conflicted with this definition and suggest that enterotypes may be less discrete.

● **Gnotobiotics** is the study of animals living in a microbiologically defined environment, either germ-free or colonized with known bacteria.

● **Inflammasomes** are protein complexes containing an intracellular sensor (such as a nucleotide-binding oligomerization domain (NOD)-like receptor), the procaspase-1 precursor and the ASC (apoptosis-associated speck-like protein containing a caspase activation and recruitment domain) adaptor protein. These complexes recognize microbe- and the host-derived inflammatory signals, microbial-associated molecular patterns and damage-associated molecular patterns, and its activation leads to the maturation of inflammatory cytokines (such as interleukin-1 β and interleukin-18). Inflammasomes participate in antimicrobial innate immune responses, but may also have a role in metabolic diseases, such as obesity, type 2 diabetes and atherosclerosis.

● **Metagenome** is the total DNA that can be extracted from an environment. The human metagenome is the aggregate of the DNA of the host and the microbiota. Metagenome and microbiome are often used interchangeably.

● **Metagenomics** is the study of the metagenome (microbiome). Metagenomics can either be targeted (usually 16S ribosomal RNA) or untargeted (shotgun sequencing).

● **Microbiota** is the collective microbial community inhabiting a specific environment. Cellular density increases along the length of the gut, and the colonic microbiota is the densest and most diverse community in the gut, and in the whole human body.

● **Microbiome** is the collective genomic content of a microbiota. It also indicates the total genetic capacity of the community.

● **Probiotics** are defined as live microorganisms that, when administered in adequate amounts, confer a health benefit for the host⁸⁴. Many bacterial strains in the *Lactobacillus* and *Bifidobacterium* genera are considered to be probiotic.

● **Prebiotics** are non-digestible food ingredients that, when consumed in sufficient amounts, selectively stimulate the growth, activity or both of one or a limited number of microbial genera or species in the gut microbiota that confer(s) health benefits to the host⁸⁵. Inulin and *trans*-galacto-oligosaccharides are defined as prebiotics because they are resistant to gastric digestion and hydrolysis by human enzymes; are fermented by specific members of the gut microbiota; and induce selective growth, activity or both of beneficial intestinal bacteria⁸⁵. Both inulin and *trans*-galacto-oligosaccharides stimulate the growth of *Bifidobacterium*, an effect defined as bifidogenic.

a gradient than discrete entities. Each enterotype is dominated by a different genus — *Bacteroides*, *Prevotella* or *Ruminococcus*²⁷ — but not affected by gender, age or nationality²⁷. Enterotypes dominated by *Bacteroides* or *Prevotella* are associated with the consumption of a diet rich in protein and animal fat, or carbohydrates, respectively²⁸. The *Ruminococcus* enterotype is not well separated and is partly merged with the *Bacteroides* enterotype²⁸. This division supports the association between *Prevotella* and a diet high in carbohydrates, which was seen in children from rural Africa²⁶. A 10-day dietary intervention, however, was not sufficient to alter the enterotype of an individual²⁸,

BOX 2

Dominant microbes

The human gut microbiota is dominated by five bacterial phyla (Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria and Verrucomicrobia) and one Archaea (Euryarchaeota). The less prevalent bacterial groups are distributed among Cyanobacteria, Fusobacteria, Lentisphaerae, Spirochaetes and TM7.

The Firmicutes phylum contains relevant genera, including *Ruminococcus*, *Clostridium*, *Lactobacillus* (several strains of which are probiotics), and the butyrate producers *Eubacterium*, *Faecalibacterium* and *Roseburia*.

In Bacteroidetes, *Bacteroides*, *Prevotella* and *Xylanibacter* degrade a variety of complex glycans.

The Actinobacteria phylum includes *Collinsella* and *Bifidobacterium* (which contains probiotic strains). Common Proteobacteria are *Escherichia* (from the Enterobacteriaceae family) and *Desulfovibrio* (which contains sulphate-reducing bacteria). Verrucomicrobia was recently discovered and includes *Akkermansia* (which are specialized for mucus degradation). Euryarchaeota contains the prevalent *Methanobrevibacter* (which is involved in the continuation of intestinal methanogenesis).

suggesting that a long-term change may be required to provoke a major shift in gut microbiota composition.

Changes in daily carbohydrate intake may affect specific groups of colonic bacteria over a short period of time. Consumption of the prebiotic inulin increases the levels of *F. prausnitzii* and *Bifidobacterium* sp. in humans²⁹. Similarly, prebiotics promote a selective increase in *Bifidobacterium* sp. in diet-induced obese mice, and this increase is correlated with reduced adiposity and levels of microbe-derived inflammatory molecules, such as lipopolysaccharide, compared with mice that are fed a high-fat diet without prebiotics³⁰. Human diets that are supplemented with resistant starch have increased faecal levels of *Ruminococcus bromii* and *Eubacterium rectale*, which correlates with fibre fermentation³¹. Consumption of resistant starch also improves insulin sensitivity³², but the variation in the microbial response to changes in resistant starch between individuals suggests successful dietary interventions need to be personalized³¹.

The gut microbiota also reacts to dietary fat. Mice fed on high-fat diets have reduced numbers of Bacteroidetes, and increased numbers of Firmicutes and Proteobacteria^{33,34}. This change is rapid, occurring within 24 hours³⁵. Transplantation of the caecal microbiota from obese mice fed on high-fat diets into germ-free recipients increases adiposity significantly more than transplantation of a lean microbiota³⁴. The altered microbial community of obese mice seems to

have some role in promoting diet-induced obesity, but the mechanisms that cause this are unknown. A change in diet clearly alters the gut microbiota, and these alterations may contribute to the host's metabolic phenotype. Further metatranscriptomic and proteomic studies should provide insight into the response of microbial function as a result of a dietary shift.

Microbial processing of food constituents

Products of microbial metabolism act as signalling molecules and influence the host's metabolism. Microbial products directly affect intestinal function but may also affect the liver and brain, as well as adipose and muscle tissue, which consequently may affect the level of obesity and the associated comorbidities (Fig. 2). Microbial enzymatic activities can act directly on the fermentation of polysaccharides and bile-acid metabolism, or act in conjunction with the host on the metabolism of choline (Fig. 3).

Fermentation of polysaccharides

Non-digestible carbohydrates are important sources of energy for several members of the colonic microbiota. Species such as *B. thetaiotaomicon* and *Bacteroides ovatus* contain more than twice the number of glycosidase and lyase genes than the human genome and are capable of using nearly all of the main plant and host glycans (such as mucus-associated glycoproteins)^{36,37}. Of the SCFAs produced from microbial fermentation, butyrate is particularly important as an energy substrate for cellular metabolism in the colonic epithelium. The colonic epithelial cells of germ-free mice are severely energy-deprived and are characterized by increased activation of AMP-activated protein kinase (AMPK), which senses cellular energy status³⁸. This is also true of the liver of germ-free mice³⁹. The liver metabolism of germ-free and colonized mice differs considerably, possibly because of the increased influx of SCFAs into the liver of colonized mice (Fig. 1). Acetate and propionate are taken up by the liver and used as substrates for lipogenesis and gluconeogenesis. Colonized mice have higher levels of stored triglycerides in the liver and an increase in the synthesis of very-low-density lipoproteins⁴⁰, which transport triglycerides from the liver to other tissues. Increased triglyceride production in the liver of colonized mice is associated with reduced expression of fasting-induced adipose factor, or ANGPTL4, in the small intestine^{4,39}. ANGPTL4 is a potent inhibitor of the enzyme lipoprotein lipase, which mediates cellular uptake of triglycerides. Germ-free *Angptl4*-deficient mice gained as much fat mass and body weight during high-fat feeding as colonized mice, indicating that ANGPTL4 directly mediates microbial regulation of adiposity in mice^{4,39}.

SCFAs also affect proliferation, differentiation and modulation of gene expression in mammalian colonic epithelial cells⁴¹. However, these effects have been attributed to butyrate acting as a potent histone deacetylase inhibitor and, as such, it may regulate 2% of the

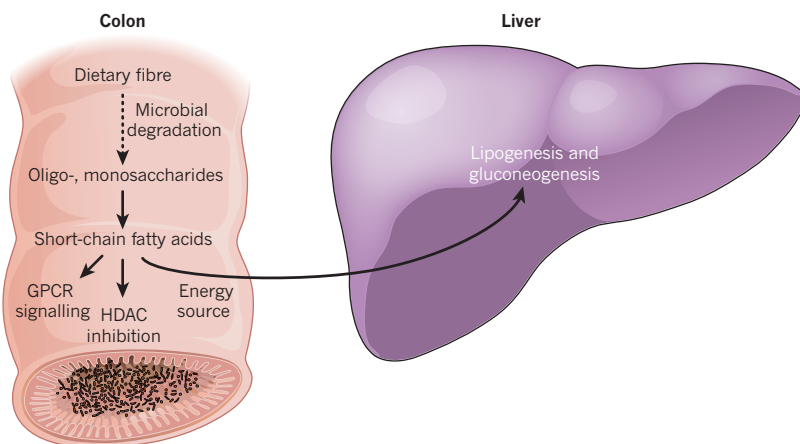


Figure 1 | Effects of colonic fermentation of dietary fibres.

Complex carbohydrates, such as dietary fibre, are metabolized by the colonic microbiota to oligosaccharides and monosaccharides and then fermented to short-chain fatty acid end-products, mainly acetate, propionate and butyrate. Short-chain fatty acids are absorbed in the colon, where butyrate provides energy for colonic epithelial cells, and acetate and propionate reach the liver and peripheral organs, where they are substrates for gluconeogenesis and lipogenesis. In addition to being energy sources, short-chain fatty acids control colonic gene expression by inhibiting the enzyme histone deacetylase (HDAC) and metabolic regulation by signalling through G-protein-coupled receptors (GPCRs), such as GPR41 or GPR43.

mammalian transcriptome⁴¹. In addition, SCFAs can regulate gene expression by binding to the G-protein-coupled receptors (GPCRs) GPR41 (also known as FFAR3) and GPR43 (also known as FFAR2). Signalling through these receptors affects several different functions depending on the cellular type. For example, SCFAs suppress inflammation through GPR43 signalling in immune cells, such as neutrophils^{42,43}, and modulate secretion of the hormone GLP-1 — which improves insulin secretion and has antidiabetic effects — by enteroendocrine L-cells in the distal small intestine and colon⁴⁴. In addition, the gut microbiota induces *Pyy* expression by L-cells through a GPR41-dependent mechanism. Conventional *Gpr41*-deficient mice have reduced adiposity compared with conventional wild-type mice, whereas germ-free wild-type and *Gpr41*-deficient mice had similar adiposity⁴⁵, indicating that the effect of the microbiota on fat deposition is dependent on this SCFA receptor.

Microbial fermentation of polysaccharides may affect host adiposity through several complementary mechanisms, so modulation of the microbiota and its fermentation capacity may provide new avenues for managing obesity.

Microbial regulation of bile-acid metabolism

The primary bile acids cholic acid and chenodeoxycholic acid are synthesized in the human liver from cholesterol, and are important for ensuring that cholesterol, dietary fats and fat-soluble vitamins from the small intestine are soluble and absorbable. Primary bile acids are conjugated to taurine in mice and to glycine in humans, and are taken up in the distal ileum for transport to the liver. However, bacteria in this part of the ileum deconjugate these bile acids, which then escape intestinal uptake and can be further metabolized by the gut microbiota into secondary bile acids. Because the gut microbiota transforms bile acids, germ-free rodents have more bile acid and a less diverse profile than their conventionally raised counterparts^{46–48}.

Bile acids also function as signalling molecules and bind to cellular receptors⁴⁹, such as the bile-acid-synthesis controlling nuclear receptor farnesoid X receptor (FXR)⁵⁰ and the GPCR TGR5. Both FXR and TGR5 have been implicated in the modulation of glucose metabolism in mice, but FXR impairs, whereas TGR5 promotes, glucose homeostasis^{51–53}. In contrast to FXR, which is activated by primary bile acids, TGR5 binds secondary bile acids such as deoxycholic acid (formed from cholic acid) and lithocholic acid (formed from chenodeoxycholic acid). TGR5 signalling in enteroendocrine L-cells induces secretion of GLP-1, thereby improving liver and pancreatic function and enhancing glucose tolerance in obese mice⁵³. Bile acids are taken up from the gut and circulated throughout the body, so activation of TGR5 and FXR in peripheral organs may contribute to overall host metabolism. Activation of TGR5 in brown adipose tissue and muscle increases energy expenditure and protects against diet-induced obesity⁵⁴. The gut microbiota may therefore contribute to the level of obesity and type 2 diabetes by controlling lipid and glucose metabolism through the composition of bile-acid pools and the modulation of FXR and TGR5 signalling.

Microbial metabolism of choline

Choline is an important component of cell membranes and is mostly obtained from foods such as red meat and eggs, but may also be synthesized by the host⁵⁵. Choline is also important for lipid metabolism and synthesis of very-low-density lipoprotein in the liver, and insufficient levels in the diet are associated with altered gut microbial ecology and liver steatosis in mice⁵⁶ and humans⁵⁷. In particular, low quantities of Gamma-proteobacteria and high levels of *Erysipelotrichi* in human faecal microbiota are associated with hepatic steatosis⁵⁷. Microbial and host enzymatic activities interact in choline's transformation into toxic methylamines, so trimethylamine that is produced by intestinal microbes can be further metabolized to trimethylamine-N-oxide in the liver^{58,59}. These transformations may decrease the levels of bioavailable choline and are suggested to trigger non-alcoholic fatty liver

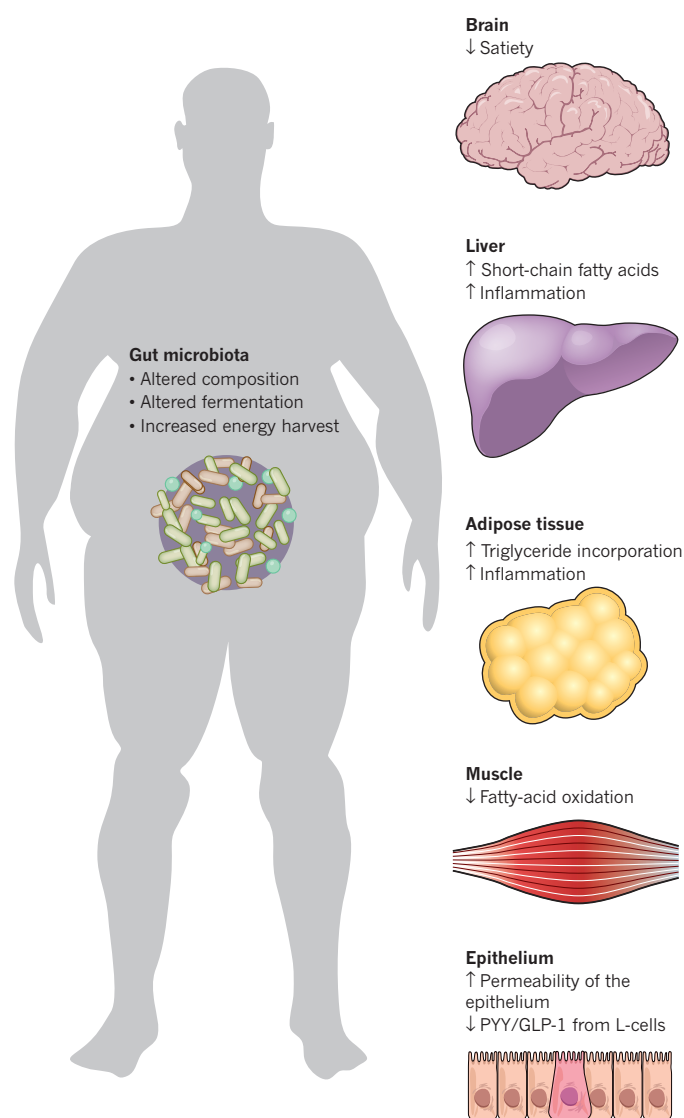


Figure 2 | Features of the gut microbiota that promote obesity and insulin resistance. Alterations to the composition and metabolic capacity of gut microbiota in obesity promote adiposity and influence metabolic processes in peripheral organs, such as the control of satiety in the brain; the release of hormones from the gut (shown as PYY and GLP-1); and the synthesis, storage or metabolism of lipids in the adipose tissue, liver and muscle. Microbial molecules also increase intestinal permeability, leading to systemic inflammation and insulin resistance.

disease (NAFLD) in mice⁵⁸. An altered gut microbial composition and its capacity to metabolize choline may have an important role in modulating NAFLD as well as glucose homeostasis⁵⁸.

Furthermore, plasma levels of trimethylamine-N-oxide and its metabolites are correlated with cardiovascular disease⁶⁰ (Fig. 3). The effect of microbial choline metabolism in cardiovascular disease is shown by the reduction of atherosclerosis in atherosclerosis-prone *Apoe*^{−/−} mice treated with broad-spectrum antibiotics⁶⁰. A gut microbiome with different capacities to process cholesterol and choline may contribute to the development of cardiovascular diseases.

Regulation of permeability and inflammation

Obesity, insulin resistance and development of type 2 diabetes are associated with systemic and adipose tissue inflammation⁶¹. The gut microbiota is a rich source of molecules such as lipopolysaccharide and peptidoglycan that may cause inflammation in peripheral tissues of the body. Colonization of germ-free mice with *Escherichia coli* is

sufficient to augment macrophage infiltration of adipose tissue and polarize macrophages towards the expression of pro-inflammatory cytokines¹⁹. Plasma lipopolysaccharide levels increase in patients with type 2 diabetes⁶², and feeding lipopolysaccharide to mice for 4 weeks increase adipose tissue inflammation and reduce insulin sensitivity⁶³. These findings suggest that the gut microbiota may affect host metabolism by altering adipose tissue inflammation. Higher numbers of T cells^{64,65} and mast cells⁶⁶, and lower numbers of regulatory T cells⁶⁷ are also involved, but if and how the gut microbiota affects these cells and whether such interactions contribute to metabolic abnormalities is unclear.

Plasma lipopolysaccharide levels seem to rise with higher fat intake in mice⁶³ and humans^{68,69}. Two hypotheses have been made to explain the mechanism: lipopolysaccharide is taken up with dietary fats in chylomicrons⁷⁰, or lipopolysaccharide reaches the circulation because the gut is more permeable in obese mice⁶³. A connection between metabolism and the function of the epithelial barrier is thought to exist. Targeted deletion of fatty acid synthase — encoded by the *Fas* gene — in the gut epithelium of mice showed that epithelial *de novo* lipogenesis is required to maintain barrier function⁷¹. *Fas*-deficient epithelium has increased permeability and, as a result, increased colonic levels of proinflammatory cytokines and high serum lipopolysaccharide. These phenotypes were corrected by antibiotic treatment, suggesting a reciprocal interaction between microbiota, altered epithelial permeability and host metabolism.

A similar connection between gut permeability and type 2 diabetes in humans could also be present. Permeability is correlated with increased visceral adiposity and hepatic steatosis⁷², and those with high visceral adiposity and type 2 diabetes have increased levels of bacterial DNA in their blood⁷³. However, inflammation may increase permeability in the gut, and further investigation into whether increased permeability causes adipose inflammation or increased inflammation contributes to increased permeability is needed. Either way, the gut microbiota modulates permeability that may contribute to adipose inflammation and cause insulin resistance.

Lipopolysaccharide molecules bind to Toll-like receptor 4 (TLR4), and peptidoglycan to nucleotide-binding oligomerization domain (NOD) receptors, both of which activates proinflammatory signalling cascades^{63,74,75}. Deletion of TLR4 in haematopoietic cells by generating bone-marrow chimaeras shows that TLR4 activation in macrophages of mice fed a high-fat diet is required for the development of fasting hyperinsulinaemia, and insulin resistance in liver and adipose tissue but not for the development of obesity⁷⁶. The innate immune system, however, also modulates microbial composition, which may have autonomous effects on host metabolism. Mice deficient in TLR5 have an altered microbial ecology and exhibit metabolic-syndrome signs, such as obesity, insulin resistance and dyslipidaemia, which are, in part, associated with increased food consumption⁷⁷. Transplantation of the gut microbiota from *Tlr5*-deficient and wild-type mice into germ-free recipients shows that the phenotypes are transmissible, and suggests that the gut microbiota alone can mediate disease.

Microbe-associated molecular patterns, including lipopolysaccharide and peptidoglycan, can be recognized by nucleotide-binding domain and leucine-rich-repeat-containing proteins (NLRPs), which form the inflammasome complex together with the apoptosis-associated speck-like protein containing a caspase activation and recruitment domain (ASC)⁷⁸. Obesity is associated with increased adipose expression of NLRP3 in mice and ablation of NLRP3 enhances insulin signalling⁷⁹. However, inflammasomes may be linked to gut microbiota and host metabolism (Fig. 4). NLRP3, NLRP6 and ASC are important regulators of microbial ecology in mice, and deletion of their genes increases the number of Bacteroidetes (Prevotellaceae) and TM7^{56,80}. In particular, deficiency in *Nlrp6* results in altered gut microbial ecology that predisposes mice to colitis⁸⁰, and inflammasome complexes that do not contain NLRP3 or NLRP6 are associated with an altered gut microbiota and promote NAFLD and non-alcoholic steatohepatitis (NASH)⁵⁶. Importantly, wild-type mice housed with disease-prone *Asc*^{-/-} mice develop NAFLD or NASH, providing direct evidence that an altered gut microbiota may cause these diseases⁵⁶. Alterations to gut microbiota composition are associated with an increased influx of TLR4 and TLR9 ligands — presumably lipopolysaccharide and bacterial DNA — respectively,

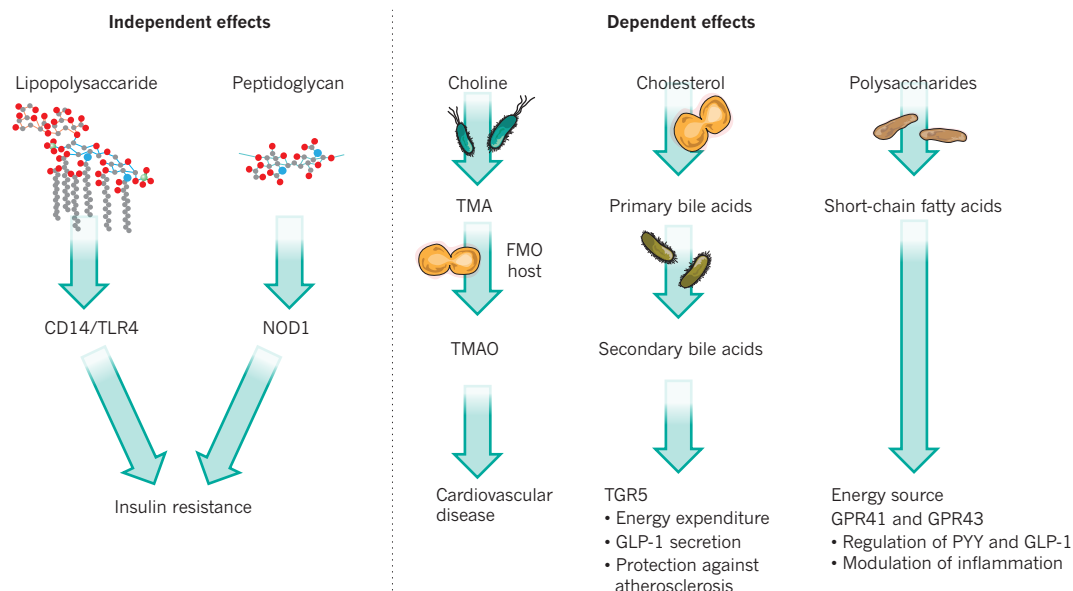


Figure 3 | Diet-independent and -dependent microbial effects on host metabolism. The gut microbiota produces pro-inflammatory molecules, such as lipopolysaccharide and peptidoglycan, which may affect host metabolism through proteins produced by the host to mediate the immune response. Choline, cholesterol and polysaccharides obtained from the diet are metabolized by the gut microbiota and either directly or through further host–microbial co-metabolization generate bioactive

compounds. In the case of choline, this can lead to cardiovascular disease; for cholesterol, activation of TGR5 can increase energy expenditure and GLP-1 secretion or protection against heart disease; and for polysaccharides, short-chain fatty acids can be used as an energy source or can bind to GPR41 or GPR43 to regulate hormones and modulate inflammation. FMO, flavin-containing monooxygenase; TLR4, Toll-like receptor 4; TMA, trimethylamine; TMAO, trimethylamine-*N*-oxide.

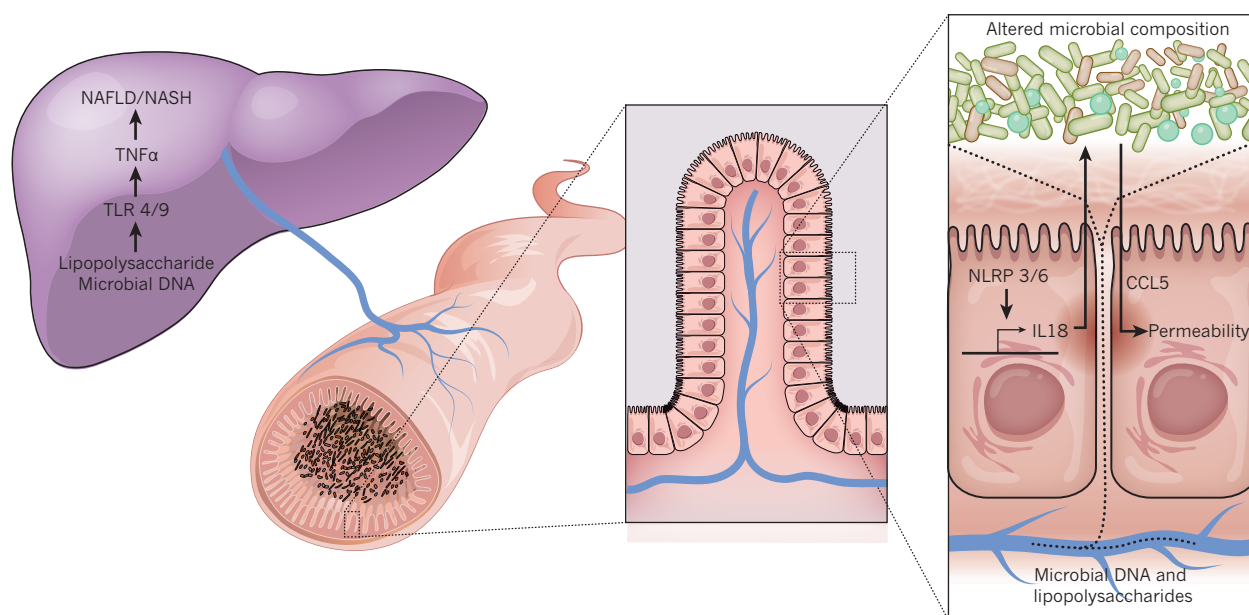


Figure 4 | Different microbial innate immune mechanisms affect host metabolism in the gut and liver. Plasma lipopolysaccharide seems to rise with higher fat intake, and those with high visceral adiposity have higher levels of microbial DNA in their blood. Both lipopolysaccharide and microbial DNA seem to be connected with gut permeability. NLRP3 and 6 are both important regulators of microbial ecology through the effector protein interleukin-18

(IL-18). The altered gut microbiota can stimulate CCL5 secretion, which can result in increased permeability and influx of microbial components. In the liver, lipopolysaccharide and bacterial DNA activate the receptors, TLR4 and 9, leading to increased tumour-necrosis-factor- α (TNF α) secretion and development of non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH).

to the liver through the portal vein⁵⁶. Mice deficient in TLR signalling in the liver are therefore protected from developing conditions related to metabolic syndrome such as obesity, NAFLD and NASH⁵⁶ (Fig. 4). The interaction between diet, host and gut microbiota may modulate gut permeability that leads to an influx of proinflammatory molecules and subsequent activation of inflammatory signalling pathways in peripheral tissues that may cause obesity, steatosis and insulin resistance.

Future research

The gut microbiota is increasingly being accepted as an environmental factor that affects host metabolism and contributes to associated pathological conditions, such as obesity, diabetes and cardiovascular disease. However, the contribution that the gut microbiota makes to causing obesity and diabetes in humans is unclear. This is probably because the heterogeneous aetiology of obesity and diabetes can be associated with different microbes; studies are underpowered and include participants with diverse ethnic origin and food habits; the composition of the gut microbiota has large interpersonal variation; and different methods, with specific biases, have been used to profile the microbiota. Cheaper sequencing and improved bioinformatics tools for the analysis of the gut microbiota will allow more researchers to use metagenomic sequencing and avoid primer and polymerase chain reaction biases linked to 16S rRNA gene surveys. Although useful, metagenomic approaches should be complemented by metatranscriptomics and metaproteomics to assess which microbial genes and proteins are expressed in specific conditions. One of the main challenges is to obtain robust predictive biomarkers for obesity and diabetes on the basis of the gut microbiota, for which improved study designs and analytical methods are essential. Much of the focus has been on the faecal microbiota, but many metabolic functions also occur in the small intestine. Sampling intestinal specimens might contribute to the identification of microbial biomarkers for health and disease, and although challenging, more emphasis should be placed on examining its microbiota and the effects on host metabolism.

Studies in humans tend to be correlative, so the role of the microbiota in obesity and its comorbidities in humans remains to be proven. However, this role can be examined in animal studies. Germ-free mice can be

'humanized' by colonizing them with human intestinal communities, providing tools for examining the function of a specific human microbiota and testing how it interacts with specific diets. Genetically engineered germ-free mice could help to identify the molecular mechanisms by which the gut microbiota affects host metabolism. Pigs have similar gastrointestinal tracts and diets to humans, so they could be useful animal models in which to test dietary interventions and to manipulate the gut microbiota to improve health and prevent disease.

Accumulating evidence indicates that the gut microbiota may be a target for treating metabolic diseases⁸¹. Supplementing the diet with non-digestible food ingredients, or prebiotics, that stimulate the expansion of specific microbes to improve metabolic regulation can be a therapy. Probiotics may be an interesting approach for prevention of obesity and related diseases. But to determine the effects of both of these therapies, double-blind, placebo-controlled studies are required. Therapies that replace unhealthy with a healthy microbiota through transplantation have been used successfully since 1958 for the treatment of antibiotic-related diarrhoeal colitis⁸². Recently, transplantation of healthy lean microbiota improved insulin signalling in participants with metabolic syndrome⁸³. Although a promising technique, transmission of unknown and potentially pathogenic bacteria and viruses from an unfractionated gut microbiota may have risks for the recipient. Using microbiota-based interventions to treat obesity will require probiotics that are selected for specific clinical manifestations of metabolic syndrome. ■

1. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010). **This article is the first catalogue of the human metagenome.**
2. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457**, 480–484 (2009). **This paper reports that the diversity of the gut microbiota differs between lean and obese individuals.**
3. Sjogren, K. *et al.* The gut microbiota regulates bone mass in mice. *J. Bone Miner. Res.* **27**, 1357–1367 (2012).
4. Backhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *Proc. Natl Acad. Sci. USA* **101**, 15718–15723 (2004). **This article reports a link between the gut microbiota and adiposity.**
5. Reinhardt, C. *et al.* Tissue factor and PAR1 promote microbiota-induced intestinal vascular remodelling. *Nature* **483**, 627–631 (2012).

6. Stappenbeck, T. S., Hooper, L. V. & Gordon, J. I. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc. Natl Acad. Sci. USA* **99**, 15451–15455 (2002).
7. Larsson, E. *et al.* Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. *Gut* **61**, 1124–1131 (2011).
8. Gaboriau-Routhiau, V. *et al.* The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* **31**, 677–689 (2009).
9. Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
10. Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006).
11. Furet, J. P. *et al.* Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes* **59**, 3049–3057 (2010).
12. Duncan, S. H. *et al.* Human colonic microbiota associated with diet, obesity and weight loss. *Int. J. Obes. (Lond.)* **32**, 1720–1724 (2008).
13. Schwartz, A. *et al.* Microbiota and SCFA in lean and overweight healthy subjects. *Obesity* **18**, 190–195 (2010).
14. Zhang, H. *et al.* Human gut microbiota in obesity and after gastric bypass. *Proc. Natl Acad. Sci. USA* **106**, 2365–2370 (2009).
15. Sjostrom, L. *et al.* Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery. *N. Engl. J. Med.* **351**, 2683–2693 (2004).
16. Sjostrom, L. *et al.* Effects of bariatric surgery on mortality in Swedish obese subjects. *N. Engl. J. Med.* **357**, 741–752 (2007).
17. Cani, P. D. *et al.* Changes in gut microbiota control metabolic endotoxemia-induced inflammation in high-fat diet-induced obesity and diabetes in mice. *Diabetes* **57**, 1470–1481 (2008).
18. Membrez, M. *et al.* Gut microbiota modulation with norfloxacin and ampicillin enhances glucose tolerance in mice. *FASEB J.* **22**, 2416–2426 (2008).
19. Caesar, R. *et al.* Gut-derived lipopolysaccharide augments adipose macrophage accumulation but is not essential for impaired glucose or insulin tolerance in mice. *Gut* <http://dx.doi.org/10.1136/gutjnl-2011-301689> (25 April, 2012).
20. Bergman, E. N. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiol. Rev.* **70**, 567–590 (1990).
21. Samuel, B. S. & Gordon, J. I. A humanized gnotobiotic mouse model of host–archaeal–bacterial mutualism. *Proc. Natl Acad. Sci. USA* **103**, 10011–10016 (2006).
22. Hoverstad, T. & Midtvedt, T. Short-chain fatty acids in germfree mice and rats. *J. Nutr.* **116**, 1772–1776 (1986).
23. Wostmann, B. S., Larkin, C., Moriarty, A. & Bruckner-Kardoss, E. Dietary intake, energy metabolism, and excretory losses of adult male germfree Wistar rats. *Lab. Anim. Sci.* **33**, 46–50 (1983).
24. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
This article demonstrates that the microbiota of those who are obese is enriched in genes for energy harvest and that obesity can be transmitted by transplantation of microbiota.
25. Nair, S., Cope, K., Risby, T. H. & Diehl, A. M. Obesity and female gender increase breath ethanol concentration: potential implications for the pathogenesis of nonalcoholic steatohepatitis. *Am. J. Gastroenterol.* **96**, 1200–1204 (2001).
26. De Filippo, C. *et al.* Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc. Natl Acad. Sci. USA* **107**, 14691–14696 (2010).
This article demonstrates the affect of diet in shaping the gut microbiota.
27. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
28. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
29. Ramirez-Farias, C. *et al.* Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. *Br. J. Nutr.* **101**, 541–550 (2009).
30. Cani, P. D. *et al.* Selective increases of bifidobacteria in gut microflora improve high-fat-diet-induced diabetes in mice through a mechanism associated with endotoxaemia. *Diabetologia* **50**, 2374–2383 (2007).
31. Walker, A. W. *et al.* Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J.* **5**, 220–230 (2010).
32. Robertson, M. D., Bickerton, A. S., Dennis, A. L., Vidal, H. & Frayn, K. N. Insulin-sensitizing effects of dietary resistant starch and effects on skeletal muscle and adipose tissue metabolism. *Am. J. Clin. Nutr.* **82**, 559–567 (2005).
33. Hildebrandt, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716–1724 (2009).
34. Turnbaugh, P. J., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
35. Turnbaugh, P. J. *et al.* The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* **1**, 6ra14 (2009).
36. Xu, J. *et al.* A genomic view of the human–*Bacteroides thetaiotaomicron* symbiosis. *Science* **299**, 2074–2076 (2003).
37. Martens, E. C. *et al.* Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol.* **9**, e1001221 (2011).
38. Donohoe, D. R. *et al.* The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab.* **13**, 517–526 (2011).
39. Backhed, F., Manchester, J. K., Semenkovich, C. F. & Gordon, J. I. Mechanisms underlying the resistance to diet-induced obesity in germ-free mice. *Proc. Natl Acad. Sci. USA* **104**, 979–984 (2007).
40. Velagapudi, V. R. *et al.* The gut microbiota modulates host energy and lipid metabolism in mice. *J. Lipid Res.* **51**, 1101–1112 (2010).
41. Davie, J. R. Inhibition of histone deacetylase activity by butyrate. *J. Nutr.* **133**, 2485S–2493S (2003).
42. Maslowski, K. M. *et al.* Regulation of inflammatory responses by gut microbiota and chemoattractant receptor GPR43. *Nature* **461**, 1282–1286 (2009).
43. Sina, C. *et al.* G protein-coupled receptor 43 is essential for neutrophil recruitment during intestinal inflammation. *J. Immunol.* **183**, 7514–7522 (2009).
44. Tolhurst, G. *et al.* Short-chain fatty acids stimulate glucagon-like peptide-1 secretion via the G-protein-coupled receptor FFAR2. *Diabetes* **61**, 364–371 (2012).
45. Samuel, B. S. *et al.* Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. *Proc. Natl Acad. Sci. USA* **105**, 16767–16772 (2008).
46. Wostmann, B. S. Intestinal bile acids and cholesterol absorption in the germfree rat. *J. Nutr.* **103**, 982–990 (1973).
47. Swann, J. R. *et al.* Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc. Natl Acad. Sci. USA* **108**, 4523–4530 (2010).
48. Gustafsson, B. E., Bergstrom, S., Lindstedt, S. & Norman, A. Turnover and nature of fecal bile acids in germfree and infected rats fed cholic acid-24-14C; bile acids and steroids 41. *Proc. Soc. Exp. Biol. Med.* **94**, 467–471 (1957).
49. Hylemon, P. B. *et al.* Bile acids as regulatory molecules. *J. Lipid Res.* **50**, 1509–1520 (2009).
50. Sinal, C. J. *et al.* Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. *Cell* **102**, 731–744 (2000).
51. van Dijk, T. H. *et al.* An increased flux through the glucose 6-phosphate pool in enterocytes delays glucose absorption in *Fxr*^{-/-} mice. *J. Biol. Chem.* **284**, 10315–10323 (2009).
52. Prawitt, J. *et al.* Farnesoid X receptor deficiency improves glucose homeostasis in mouse models of obesity. *Diabetes* **60**, 1861–1871 (2011).
53. Thomas, C. *et al.* TGR5-mediated bile acid sensing controls glucose homeostasis. *Cell Metab.* **10**, 167–177 (2009).
54. Watanabe, M. *et al.* Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **439**, 484–489 (2006).
55. Vance, D. E. Role of phosphatidylcholine biosynthesis in the regulation of lipoprotein homeostasis. *Curr. Opin. Lipidol.* **19**, 229–234 (2008).
56. Henao-Mejia, J. *et al.* Inflammasome-mediated dysbiosis regulates progression of NAFLD and obesity. *Nature* **482**, 179–185 (2012).
This paper elegantly connects microbial recognition by inflammasomes in the gut with altered permeability and liver steatosis.
57. Spencer, M. D. *et al.* Association between composition of the human gastrointestinal microbiome and development of fatty liver with choline deficiency. *Gastroenterology* **140**, 976–986 (2011).
58. Dumas, M. E. *et al.* Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc. Natl Acad. Sci. USA* **103**, 12511–12516 (2006).
59. Prentiss, P. G. *et al.* The metabolism of choline by the germfree rat. *Arch. Biochem. Biophys.* **94**, 424–429 (1961).
60. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
61. Osborn, O. & Olefsky, J. M. The cellular and signaling networks linking the immune system and metabolism in disease. *Nature Med.* **18**, 363–374 (2012).
62. Creely, S. J. *et al.* Lipopolysaccharide activates an innate immune system response in human adipose tissue in obesity and type 2 diabetes. *Am. J. Physiol. Endocrinol. Metab.* **292**, E740–E747 (2007).
63. Cani, P. D. *et al.* Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56**, 1761–1772 (2007).
This paper reports that endotoxin is sufficient to alter metabolic inflammation and insulin sensitivity.
64. Nishimura, S. *et al.* CD8⁺ effector T cells contribute to macrophage recruitment and adipose tissue inflammation in obesity. *Nature Med.* **15**, 914–920 (2009).
65. Winer, S. *et al.* Normalization of obesity-associated insulin resistance through immunotherapy. *Nature Med.* **15**, 921–929 (2009).
66. Liu, J. *et al.* Genetic deficiency and pharmacological stabilization of mast cells reduce diet-induced obesity and diabetes in mice. *Nature Med.* **15**, 940–945 (2009).
67. Feuerer, M. *et al.* Lean, but not obese, fat is enriched for a unique population of regulatory T cells that affect metabolic parameters. *Nature Med.* **15**, 930–939 (2009).
68. Amar, J. *et al.* Energy intake is associated with endotoxemia in apparently healthy men. *Am. J. Clin. Nutr.* **87**, 1219–1223 (2008).
69. Erridge, C., Attina, T., Spickett, C. M. & Webb, D. J. A high-fat meal induces low-grade endotoxemia: evidence of a novel mechanism of postprandial inflammation. *Am. J. Clin. Nutr.* **86**, 1286–1292 (2007).

70. Ghoshal, S., Witta, J., Zhong, J., de Villiers, W. & Eckhardt, E. Chylomicrons promote intestinal absorption of lipopolysaccharides. *J. Lipid Res.* **50**, 90–97 (2009).
71. Wei, X. *et al.* Fatty acid synthase modulates intestinal barrier function through palmitoylation of mucin 2. *Cell Host Microbe* **11**, 140–152 (2012).
72. Gummesson, A. *et al.* Intestinal permeability is associated with visceral adiposity in healthy women. *Obesity* **19**, 2280–2282 (2011).
73. Amar, J. *et al.* Involvement of tissue bacteria in the onset of diabetes in humans: evidence for a concept. *Diabetologia* **54**, 3055–3061 (2011).
74. Amar, J. *et al.* Intestinal mucosal adherence and translocation of commensal bacteria at the early onset of type 2 diabetes: molecular mechanisms and probiotic treatment. *EMBO Mol. Med.* **3**, 559–572 (2011).
75. Schertzer, J. D. *et al.* NOD1 activators link innate immunity to insulin resistance. *Diabetes* **60**, 2206–2215 (2011).
76. Saberi, M. *et al.* Hematopoietic cell-specific deletion of Toll-like receptor 4 ameliorates hepatic and adipose tissue insulin resistance in high-fat-fed mice. *Cell Metab.* **10**, 419–429 (2009).
77. Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
78. Strowig, T., Henao-Mejia, J., Elinav, E. & Flavell, R. Inflammasomes in health and disease. *Nature* **481**, 278–286 (2012).
79. Vandanmagsar, B. *et al.* The NLRP3 inflammasome instigates obesity-induced inflammation and insulin resistance. *Nature Med.* **17**, 179–188 (2011).
80. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757 (2011).
81. Murphy, E. F. *et al.* Divergent metabolic outcomes arising from targeted manipulation of the gut microbiota in diet-induced obesity. *Gut* <http://dx.doi.org/10.1136/gutjnl-2011-300705> (16 February, 2012).
82. Eiseman, B., Silen, W., Bascom, G. S. & Kauvar, A. J. Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. *Surgery* **44**, 854–859 (1958).
83. Vrieze, A. *et al.* Transfer of intestinal microbiota from lean donors increases insulin sensitivity in subjects with metabolic syndrome. *Gastroenterology* <http://dx.doi.org/10.1053/j.gastro.2012.06.031> (20 June, 2012)
This article reports that insulin resistance in obese humans can be improved by transplantation of a lean microbiota.
84. Food and Agriculture Organization of the United Nations and World Health Organization. *Working Group Report on Drafting Guidelines for the evaluation of Probiotics in Food.* <ftp://ftp.fao.org/esn/food/wgreport2.pdf> (FAO and WHO, 2002).
85. Roberfroid, M. *et al.* Probiotic effects: metabolic and health benefits. *Br. J. Nutr.* **104**, S1–S63 (2010).

Acknowledgements The authors are grateful to A. Hallén for contribution to the artwork and to R. Perkins for reading the manuscript. Work in the authors' laboratory is funded by the Swedish Research Council, the Swedish Foundation for Strategic Research, Torsten Söderberg's Foundation, Ragnar Söderberg's foundation, AFA Insurances, the Knut and Alice Wallenberg foundation, the Swedish heart lung foundation, the NovoNordisk foundation and the Swedish diabetes foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/1azgex. Correspondence should be addressed to F.B. (Fredrik.Backhed@wlab.gu.se).

Genomic approaches to studying the human microbiota

George M. Weinstock¹

The human body is colonized by a vast array of microbes, which form communities of bacteria, viruses and microbial eukaryotes that are specific to each anatomical environment. Every community must be studied as a whole because many organisms have never been cultured independently, and this poses formidable challenges. The advent of next-generation DNA sequencing has allowed more sophisticated analysis and sampling of these complex systems by culture-independent methods. These methods are revealing differences in community structure between anatomical sites, between individuals, and between healthy and diseased states, and are transforming our view of human biology.

The microbes that exist in the human body are collectively known as the human microbiota. This amazingly complex and poorly understood group of communities has an enormous impact on humans. An increasing number of conditions are being examined for correlative and causative associations with the microbiome — which, in this Review, is used to refer to the microbiota and the habitat it colonizes (Box 1). Each one of the many microbial communities has its own structure and ecosystem, depending on the body environment it exists in. The fundamental goal of human microbiome research is to measure the structure and dynamics of microbial communities, the relationships between their members, what substances are produced and consumed, the interaction with the host, and differences between healthy hosts and those with disease.

Despite an explosion in human-microbiome research, these communities are still the dark matter of the body. The microbiome has been called another organ^{1–4} because of its products, its responsiveness to the environment and its integration with other systems. Sometimes referred to as our second genome⁵, the genes of microbes that make up the microbiome outnumber human genes by more than 100-fold, with over 3 million bacterial genes in the gut alone^{6,7}. These extensive microbial ecosystems are not limited to the human body. Microbes and their communities dominate the environment and occupy a vast range of niches. Environmental metagenomics was developed extensively before being applied to the human body^{8,9}, and methods from other disciplines have had a significant effect on human-microbiome research. Defining complicated microbial ecosystems and developing tools to probe their workings is an important research enterprise of twenty-first century microbiology.

The complexity of microbial communities makes studying them challenging. There may be hundreds of different species, and enumerating what organisms are present with standard microbiological techniques is not possible because many organisms have never been grown in culture and may require special, as yet unknown, growth conditions. In addition, the abundance of some microbes can range over orders of magnitude, so deep sampling is required to detect the less-abundant members. Culture-independent methods of taking a microbial census began about 25 years ago and were based on targeted sequencing of 5S and 16S ribosomal RNA genes¹⁰, which differ for each species and are a convenient identifier. As this became a tractable research area, next-generation sequencing (NGS) technologies (Table 1) were developed and allowed more extensive analyses,

both targeted 16S rRNA gene sequencing and whole-genome shotgun sequencing of microbes in communities en masse. The number of culture-independent metagenomic investigations of the human microbiome has mushroomed, and it is one of the most studied areas of microbiology with significant potential to benefit clinical practice. This culture-independent methodology is broadly applied outside human-microbiome research and is expanding our knowledge of the environment. This Review describes how NGS approaches are transforming human-microbiome studies, and posing questions and challenges for the future.

Single organisms and microbial communities

In the past, research on microbial interactions with humans has focused on single pathogenic organisms. Studies of communities of non-pathogenic microbes in the body were limited because the organisms were thought to be benign, with minor effects on human health compared with pathogens. Microbiome research has led to new interest in the communities of non-pathogenic microbes that inhabit the human body, and the need to describe the genomes of these organisms to understand the human microbiome has been recognized.

Every community of the microbiome has its own characteristics (Table 2). For the gut community, for example, high biodiversity is associated with a healthy state and reduced biodiversity occurs in patients with conditions such as Crohn's disease¹¹, whereas for tissues of the vagina, a lower biodiversity exists in healthy individuals and a bloom of organisms occurs in patients with vaginosis¹². To understand why different sites have different properties, the mechanisms that lead to the disruption of ecosystems and to disease, and exceptions to generalities about a tissue, researchers require knowledge of the structure and behaviour of microbial communities.

Microbial communities benefit the host by providing functions such as digestion of nutrients¹³ or protection against infection¹⁴. Antibiotic treatment perturbs the microbiome^{15,16} by reducing its size and altering its composition. This disturbance can lead to infection^{17–19}, and antibiotic-resistant organisms such as *Clostridium difficile* — normally controlled by the microbiome — can overgrow and create problems²⁰. More complex community contributions also exist, such as interactions with host immune and inflammatory systems^{21,22} or production of metabolites involving hybrid pathways from multiple organisms, including host-microbe pathways²³. Understanding these phenomena will ultimately allow the

¹The Genome Institute, Washington University, 4444 Forest Park Avenue, Campus Box 8501, St. Louis, Missouri 63108, USA.

microbiome to be manipulated so that, for example, transplants of microbial communities could treat *C. difficile* infections^{24,25}.

Whether the microbial ecology of the human body can be simplified to the properties of single organisms is unknown. Many organisms have never been cultured and may be adapted to life in a community environment rather than a pure culture. For organisms for which growth requirements are understood, there is a dependence on secreted products from other community members. For example, secreted siderophores²⁶ are small molecules that help microbes to scavenge iron, which is a limiting factor for growth in the body. So even the study of individual organisms can be dependent on studying the community.

Dissecting a microbiome

Analysis of community structure (Fig. 1) focuses on either targeted regions (such as the 16S rRNA gene) or shotgun sequencing to catalogue the genes that are present. Additional analysis involves sequencing genomes of individual organisms to produce a catalogue of reference genomes²⁷, and analysing RNA to describe the transcriptome and identify RNA viruses. Non-genomic analyses include proteomic and metabolomic studies, but these are not discussed here. Every sample should be well-annotated with clinical metadata, so that, ultimately, the microbiome's genetic and community structures can be correlated with the individual's phenotype.

Census of organisms

Modern metagenomic analyses of microbial communities were developed from culture-independent methods for taking a census of organisms present in a community and their abundances. Although DNA reassociation kinetics provides information on community diversity and structure²⁸, there is no accounting for organisms that may be tracked between samples. Methods more useful for providing information on the entire structure often focus on signature sequences that distinguish taxa (detected by hybridization to arrays of diagnostic oligonucleotides²⁹), various methods for fingerprinting polymerase chain reaction (PCR) products (such as single-strand conformation polymorphisms or terminal restriction fragment length polymorphisms) or DNA sequencing of targeted PCR products. Sequencing of 16S rRNA genes is the main method of taking a community census because fingerprinting methods do not adequately measure low-abundance organisms³⁰.

16S rRNA differs for each bacterial species. A bacterial species is hard to define, but is often taken as organisms with 16S rRNA gene sequences having at least 97% identity — an operational taxonomic unit (OTU). A 16S rRNA gene sequence of about 1.5 kilobases has nine short hypervariable regions that distinguish bacterial taxa; the sequences of one or more of these regions are targeted in a community census.

Before the introduction of NGS methods, the prevailing approach was to clone full-length 16S rRNA genes after PCR with primers that would amplify genes from a wide range of organisms. Cloned 16S rRNA genes were sequenced by the Sanger method, which required two or three reads to cover the entire gene. Accuracy was crucial because sequencing errors led to misclassification. The cost and effort required for the Sanger method limited the depth of sampling, and studies often produced about 100 sequences per specimen. This method identified the dominant organisms in a community, but analysis of less abundant organisms was limited.

Introducing NGS to 16S rRNA gene analysis led to marked improvements in cost and depth of sampling. The Roche-454 platform has dominated microbial community analysis³¹. As the read length for 454 pyrosequencing is about 400 bases, only a portion of the 16S rRNA gene can be sampled, and many different studies have targeted between one and three of the hypervariable regions, with different hypervariable regions targeted in different studies. Using a portion of the 16S rRNA gene led to a loss of sensitivity (some taxa

BOX 1

Terminology

- **Biodiversity** is a measure of the complexity of a community. It is affected by the number of taxa (richness) and their range of abundance (evenness). High biodiversity occurs when many taxa (high richness) are present at similar abundances (an even distribution).
- **Commensals** are organisms that benefit from another organism but that have no harm or benefit themselves. Microbes of the microbiome were thought to be commensals that benefited from the human host but did no harm. Many of these organisms provide benefits to the human host and so have a mutualistic relationship.
- **Contig** is a stretch of contiguous sequence in a genome assembly.
- **Coverage** is the number of times a genome or gene is sequenced. In a genome sequenced to coverage, each nucleotide in the sequence appears, on average, in 100 reads.
- **Genome assembly** is the process of constructing a genome sequence from short subsequences by sequencing many random fragments from a sheared genome. The random short sequences are compared, and overlapping common sequences are used to determine their orientation and order with respect to each other. A consensus sequence is constructed from this layout. Usually there are gaps, but when contigs can be arranged in the correct order and orientation, these longer stretches are called scaffolds.
- **Metagenomics** was defined⁸³ as a process for identifying genes specifically by their function by cloning them directly from the environment and expressing genes in a surrogate host⁸⁴. Therefore, gene function was known even if the sequence was not sufficient for functional inference, such as when it encoded a protein of previously unknown function. This definition, also known as functional metagenomics, is widely used. More recently, metagenomics refers to general analyses of microbial communities by culture-independent methods, which do not necessarily focus on function. The combined genomes of the microbes in a community are thought of as the community metagenome. Another type of metagenomic analysis focuses on the structure of these aggregate genomes in a community.
- **Microbiome** in this Review refers to the microbiota and the habitat it colonizes and is analogous to the term biome in ecology. Microbiome is also used to refer to the collective genomes of the microbes — what is now the metagenome, and may have originally been coined by Joshua Lederberg (cited by Hooper and Gordon⁸⁵). However, it is also used for the more ecologically consistent meaning. A microbiome can be a specific body site, such as the gut microbiome, but the human microbiome is often used to refer to the collection of microbiomes of the human body.
- **Mutualism** is a type of symbiosis in which both organisms benefit. This is one type of relationship seen in the human microbiome.
- **Operational taxonomic unit** in microbiome research is a group of organisms with 16S ribosomal RNA gene sequences that show a certain level of identity. This group is often used as a surrogate for a species when the 16S rRNA sequences are at least 97% identical.
- **Pathogenic microbe** is one with the potential to cause disease.
- **Read** is the primary output of DNA sequencing, consisting of a short stretch of DNA sequence that is produced from sequencing a region of a single DNA fragment.
- **Shotgun sequencing** is the process of randomly breaking (often by shearing) a long DNA molecule (for example, a complete chromosome) and then sequencing the resultant DNA fragments, which each come from a different location in the original long DNA molecule.
- **Virome** is the collection of viruses in the microbiota.

Table 1 | DNA sequencing platforms used for microbiome analysis

Platform	Method	Characteristics	16S rRNA	Shotgun	Comments
Established					
Sanger-based or capillary-based instrument	Fluorescent, dideoxy terminator	750-base reads High accuracy	Full length sequenced with 2–3 reads	Long reads help with database comparisons	Most costly method Relatively low throughput, so low coverage of 16S or shotgun
Roche-454	Pyrosequencing light emission	400-base reads	Up to 3 variable regions per read	Long reads help with database comparisons	Cost limits shotgun coverage but 16S coverage is good
Illumina	Fluorescent, stepwise sequencing	100–150-base reads	Only 1 variable region per read	Short reads do not seem to limit analysis	Very high coverage owing to high instrument output and very low cost
Not yet widely used					
Ion Torrent	Proton detection	More than 200-base reads	Like other NGS	Like Illumina	Expect high coverage, but longer reads than Illumina
PacBio	Fluorescent, single-molecule sequencing	Up to 10-kilobase reads Low accuracy	Accuracy an issue for correct taxon identification	Long reads could help assembly	Attractive for long reads, but lower accuracy limits applications
Oxford Nanopore*	Electronic signal as DNA passes through pore Single-molecule sequencing	Long reads	Unknown	Long reads could help assembly	Not yet available

*At the time of publication, the Oxford Nanopore system was not available, and information provided is based on company presentations. Ion Torrent and PacBio are both available but have not been widely used for microbiome analysis. The Illumina MiSeq instrument is expected to provide 250-base reads in the near future.

cannot be reliably defined at the species level, although high confidence identification of higher taxonomic ranks is possible), nevertheless gains in depth of sampling and cost savings outweigh this caveat. The US Human Microbiome Project (HMP)³² has sequenced more than 10,000 specimens from healthy adults on the 454 platform by targeting V3 to V5 regions in the 16S rRNA gene and producing, on average, 7,000 sequences per specimen³³, which is a vast expansion on the Sanger method of sequencing analysis. The results of the HMP, which sampled 18 body sites, provide an in-depth definition of the human microbiome. Another study¹⁶ that focused on the effects of the antibiotic ciprofloxacin reported the ‘rare biosphere’ in the gut. This study documented perturbation of taxa and recovery from antibiotic treatment, as well as minor constituents that did not recover after antibiotic treatment. Such analyses will be important in identifying individuals who are at risk of side effects from antibiotic treatment, for example overgrowth of pathogens such as *C. difficile* or life-threatening antibiotic-associated diarrhoea.

When using 16S rRNA gene sequencing to compare individuals it is not necessary to know which organisms are present, only whether the spectra of 16S rRNA gene sequences are similar and the degree of difference between samples. Projects that compare healthy cohorts and those with disease to determine whether there is a difference in the microbiome, or examine the effects of diet, antibiotic treatment or environmental factors on the microbiome, all focus on detecting differences in communities, rather than identifying actual taxa. A loss of sensitivity for organism identification can be tolerated, and NGS allows cost-effective deep sampling of large cohorts, which is needed to reach statistically significant conclusions. The Illumina sequencing platform has been applied to metagenomics projects^{34–36}, but because this sequencing platform currently produces reads of 100 bases (HiSeq system) to 150 bases (MiSeq system), only a single hypervariable region can be sequenced. However, this further loss of sensitivity does not preclude the use of the Illumina platform for the comparative projects already described in this Review. An early application of this platform was its use in a study of vaginal microbiomes in patients with HIV, for which comparisons of patients with conditions such as vaginosis before and after antibiotic therapy were examined³⁷. As a result of the exceptional increases in numbers of reads and the lower cost associated with the Illumina platform, it is becoming more widely used for 16S rRNA gene-sequence profiling and continues the microbiome-analysis trend of deeper sampling at lower costs.

Shotgun sequencing for cataloguing organisms

Targeted sequencing is a powerful tool for assessing the organisms that are present in microbial communities, but it is limited in terms of the functional and genetic information produced. Organisms for which the genome sequences are known (currently there are several thousand sequenced bacterial genomes) can be used to infer the genes and functional capabilities of the community (Fig. 1). However, many organisms have no reference sequence. Furthermore, a reference sequence does not completely describe the genes that are contributed by an organism. There is considerable variation in the genomes between strains of the same species. Two strains of *Escherichia coli*, O157:H7 and K-12, both have 16S rRNA gene sequences of *E. coli*, but differ in hundreds of genes. There are limits to what can be learned about the genetic content of communities from 16S rRNA gene sequences alone.

Moving beyond this level of functional inference requires a gene-based census. This catalogue of genes can be provided by shotgun sequencing of DNA that has been extracted from the community as a whole and samples the mixture of genomes that make up the metagenome (Fig. 1). In a community in excess of hundreds of species with varying abundance, deep sequencing is needed to sample minor constituents that are not necessarily unimportant. The bacterial concentration in the gut can be 10^{11} cells ml⁻¹ (refs. 38, 39), so for an organism that is present at a concentration of 1 per 10^6 there are 10^5 cells ml⁻¹, which is sufficient for the organism's products, such as metabolites and toxins, to have an effect on the community and the host.

Illumina sequencing of faecal samples produced 4 gigabases per sample and 10 Gb per sample in the Metagenomics of the Human Intestinal Tract (MetaHIT)⁶ and HMP³³ projects, respectively, which corresponded to tens of millions of reads per sample. At this depth of sequencing, the genomes of minor constituents such as *E. coli* (with an abundance of about 1% or lower) are sampled almost completely, and organisms with an even lower abundance have some of their genome represented. This extraordinary sampling of complex microbial communities is made possible by producing large amounts of data and by the low cost of NGS methods.

Shotgun sequence data, in addition to 16S rRNA gene analysis, provide information on the organisms that make up communities. Extracting 16S rRNA gene sequences from shotgun reads to determine the organisms present is possible; however, targeted 16S rRNA gene sequencing tends to introduce biases (owing to the broad-range PCR used to amplify 16S rRNA gene sequences or the choice of region within the 16S rRNA gene), which shotgun sequencing does not. Shotgun sequencing is less sensitive than targeted rRNA sequencing because a small fraction of the

sequences are from 16S rRNA genes. Another approach is to align shotgun sequences to bacterial reference genomes^{33,40,41}, allowing the relative abundance of species to be determined on the basis of the number of reads that align to each reference genome (also useful for the comparative studies already described). The MetaHIT project has used this approach to classify individuals into different groups, called enterotypes, on the basis of the community structure in their faecal samples⁴⁰. The same enterotypes have been found in 16S rRNA gene-based analysis⁴². The vaginal microbiome has also been classified into five groups⁴³. These observations suggest the human microbiome may exist in distinct states in different people, although correlation with environmental, genetic or health status is not yet clear. Stratifying future studies depending on which community class an individual belongs to may be important for identifying correlations with phenotypic data.

The need for reference genome sequences is clear both to infer genetic content of organisms identified by 16S rRNA genes and to identify sources of shotgun reads by aligning to reference genomes, and so determining organismal content of communities from shotgun data. NGS techniques have reduced the cost of bacterial sequences to less than US\$1,000 per genome and led to an increase in the production of 'complete' genome sequences. Current methodology relies mainly on Illumina shotgun sequencing and a variety of methods to assemble the reads into a genome. The product is not a true complete genome, but a high-quality draft that covers almost all of the genome and results in a high-quality base sequence²⁷. Programmes such as the HMP^{32,44} and the Genomic Encyclopedia of Bacteria and Archaea (GEBA)⁴⁵ are producing reference genomes by the thousands.

Although bacteria are the main components of the human microbiome, eukaryotic microbes and viruses (both human viruses and bacteriophages) are also present (Table 2). The study of eukaryotic microbes is not as advanced as that of bacteria⁴⁶, but the organisms are identified by signature sequences (such as fingerprinting and 18S rRNA) and shotgun sequencing analogous to bacteria. The number of reference genomes for eukaryotic microbes is smaller than that for bacteria, and progress will depend on addressing this shortfall.

By contrast, considerable effort is being given to characterizing the genomes of human viruses⁴⁷ and bacteriophages⁴⁸, known as the virome (Box 1). This work is based on shotgun sequencing (Fig. 1), although oligonucleotide microarrays for virus detection are also used^{49,50}. Viral sequences can be detected in shotgun data from different body sites, and viruses can also be enriched by processing samples before DNA extraction⁵¹. Virome analysis by shotgun sequencing of microbial communities (discussed later) has led to the identification of human viruses^{52–54}, as well as the detection of known viruses in healthy subjects and diseases of unknown aetiology⁵⁵. Likewise, bacteriophages are found to be highly diverse at different body sites^{56–58}, with differences between individuals as a result of diet⁵⁹ or disease states^{60,61}.

Sequencing for gene catalogues and functional inference

Metagenomic shotgun data also sample community gene content, which is useful to define community capabilities and identify

particular members. Deep sequencing, such as that used in the MetaHIT and the HMP, broadly samples the genomes of even minor constituents, facilitating the identification of genes present within a given community (Fig. 1). By using the sequence reads themselves, or by first assembling them into contigs (Box 1), sequence data can be compared with databases such as the National Institutes of Health's GenBank to identify which genes are present. *De novo* prediction of genes from metagenomic data is also possible³³, which provides motifs for functional inference even if the sequence does not find a match in a database. Finally, alignment of reads or contigs to reference genomes identifies which organisms are present, along with their known gene content. These methods convert metagenomic sequence data into catalogues of genes that can be further analysed.

Gene catalogues can be compared with databases such as the Kyoto Encyclopedia of Genes and Genomes (KEGG)⁶², which sorts gene products into pathways and processes. Such analyses provides lists of pathways, identify which pathway genes are in the community and quantify the abundances of genes and pathways⁶³. Comparing gene catalogues to specialized metabolic databases, such as the Carbohydrate-Active Enzymes database⁶⁴, is also useful. Carbohydrate-degrading capabilities of communities differ between body sites, suggesting the carbohydrate spectrum of each body site has determined which organisms and pathways are present⁶⁵.

In addition to pathway analysis, determining the presence and abundance of genes, such as antibiotic-resistance genes or virulence factors, in a community is possible using similar methods to those already described, and can shed light on pathogen burden in an individual and consequences of antibiotic treatment. The importance of functional analyses cannot be overemphasized, and functional properties of communities are thought to be more important than their taxonomic composition⁶⁶.

Computational tools and strategies

Metagenomic data are a rich source of information for the sequencing and analysis methods already discussed^{67,68}. The data analysis workflow has three phases. In the first phase, primary data are processed and filtered depending on the application. For 16S rRNA gene sequencing, the quality of analysis is important so that organisms are not misclassified. Initial processing addresses read quality, chimerism (a read formed from different 16S rRNA genes), read length after removing low-quality bases and related issues^{69–73}. For shotgun sequence data^{6,33} — in addition to sequence quality — artefacts such as duplicate reads must also be addressed, as well as computationally removing contamination from human sequences. Removal of human and bacterial sequences is important in read processing for virome analysis^{47,55} (Fig. 1).

Following production of processed reads, the second phase involves generating various derivative data sets. For 16S rRNA gene analysis, tables of taxa and abundance are produced by comparisons with 16S rRNA sequence databases or by using software packages to

Table 2 | Characteristics of bacteria, microbial eukaryotes and viruses in the human microbiome

Characteristic	Bacteria	Viruses	Eukaryotic microbes
Genome size	0.5–10 megabases	1–1,000 kilobases	10–50 megabases
Number of taxa in the human microbiome	At least thousands	Unknown, but could be as many as bacteria	Unknown, but may be fewer than bacteria
Relative abundances	Highly variable	Highly variable	Unknown
Targeted detection methods	Sequencing of genes such as 5S and 16S rRNA	No universal method for genes, but virus-specific polymerase chain reaction assays for some	Sequencing of 18S rRNA gene Spacer region in rRNA
Shotgun approach to analyses	Alignment to reference genomes or database comparison	Database comparison	Alignment to reference genomes or database comparison
Subspecies or strain diversity	Modest sequence variation Horizontal gene transfer also contributes	High sequence variation	Unknown

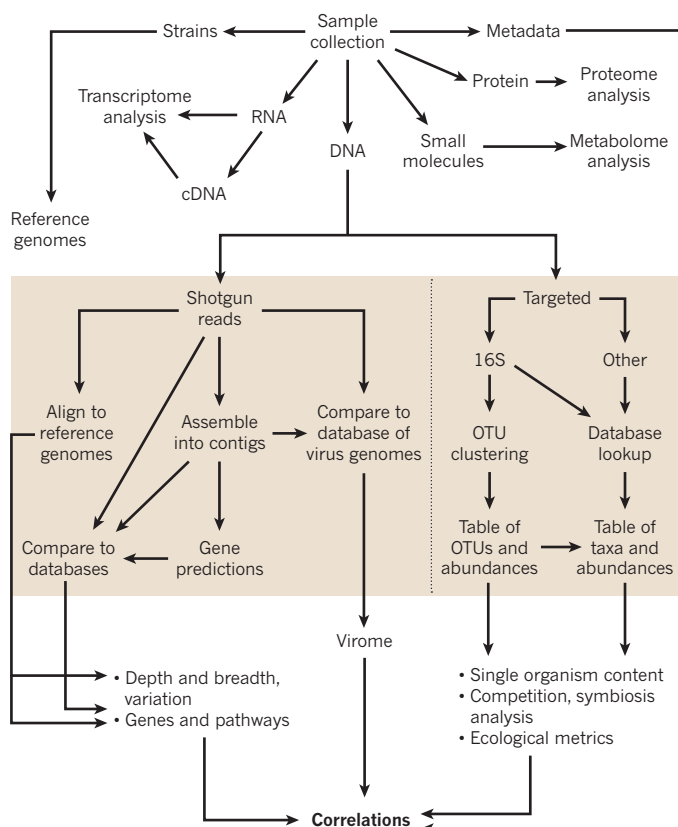


Figure 1 | Data and analysis workflow for microbiome analysis. From a microbiota sample, DNA, RNA and protein can be extracted, and metadata and strains of bacteria obtained. Data from DNA can be supplemented with proteome and transcriptome analysis. During primary analysis, shotgun techniques can produce reads from DNA, which are then aligned to reference genomes to identify variants and community population genetics, assembled into contigs to make gene predictions or compared with databases. Alternatively, targeted sequencing such as 16S rRNA gene sequencing can be used to take a community census, and these data are then compared with databases to create tables of taxa and abundance, or analysed with software programs to cluster the reads into OTUs to create tables of abundance. The derivative data is used in secondary analysis for ecological metrics or competition and symbiosis analysis. In addition, shotgun reads and comparisons with reference genomes and databases can be used to build pathways and reconstruct the capabilities of a community. The combination of these analyses will contribute to understanding the differences within and between individuals.

cluster the reads into OTUs^{74,75}. Comparing shotgun reads to gene databases, such as GenBank or KEGG, by using the Basic Local Alignment Search Tool (BLAST), for example, produces lists of genes and the number of matched reads^{7,33,63}. Alignment of reads to reference genomes produces tables of breadth and depth of coverage, by reads of each genome⁴¹. In each of these data sets, there is more biological information to be gleaned and added through further analysis. Not all reads match sequences in databases because not all organisms have a reference genome sequenced. In addition, reads may match genes whose function has not been elucidated. These sequences of unknown origin or function can be a sizeable fraction and the effect of this uninformative portion of data on analyses and conclusions is not clear.

The third phase of analysis uses these derivative data to produce trees or other representations of the similarity of communities, abundance curves, biodiversity plots, and other ecological and statistical descriptors of community structure^{74,75} (Fig. 1). A list of hits from BLAST is used to build metabolic pathways for reconstruction of community capabilities⁶³. Alignments to reference genomes are

further analysed for variants and population genetics of communities. Computational analysis can also be used to determine which organisms co-occur or rarely co-occur as evidence for symbiosis or competition, respectively, or to follow the dynamics of community structure in longitudinal time series⁷⁶.

Some analyses pose significant computational challenges. Comparisons to gene databases at the protein level are particularly demanding because shotgun sequences must be translated into polypeptides in all six reading frames, and each must be compared with a gene database represented at the protein level. Using conventional BLASTx programs for this comparison in large data sets, such as the HMP, could take decades, so supercomputers, accelerated BLAST programs or both must be used³³. A lack of efficient software and large enough computer clusters are often bottlenecks for metagenomic analysis, because sequencing and data production are not limiting factors. Management of large data sets and computing resources are receiving more attention, with cloud-computing services seeming to be a viable alternative⁷⁷.

Future directions and challenges

The rapid rise in metagenomic studies has solved many problems but, as the field has grown, other questions have been raised. Existing methodology is becoming more sophisticated, and sequencing technology is making exponential advances (Table 1). The Illumina platform introduced instruments that were more appropriate for sequencing smaller genomes, with faster run times and longer read lengths, offering more flexibility for metagenomic applications. The long read length of the PacBio platform has the potential to help distinguish the reads from different organisms, which is a challenge for metagenomic shotgun sequencing. The technology produced by Oxford Nanopore promises long reads and short run times in a scalable system, and is therefore a good match for microbial applications. Reducing the amount of DNA needed for shotgun sequencing will allow communities in smaller anatomical regions, such as within the gastrointestinal tract, to be studied separately rather than together with other regions as is the case with the current methodology. Short run-time instruments and reductions in sample size will also hasten the introduction of microbiome analysis to the clinic, where analyses of patient samples must be quick and able to deal with limited amounts of material. Ultimately, the aim of human-microbiome research is its application as a diagnostic, therapeutic and preventive tool in the clinic.

The main limitation of using shotgun data is the large number of organisms that have not been cultured, let alone sequenced. These organisms are therefore under-represented in databases, and their shotgun reads are anonymous. When community shotgun data are assembled into genomes to obtain genome sequences for new organisms, contig sizes are typically small as a result of lower organism abundance and the challenges associated with assembly of a complex mixture. The long read lengths of PacBio and Oxford Nanopore instruments should help with these challenges, as will the development of assembly algorithms for metagenomic data. Expanding the catalogue of reference genomes by producing reference sequences for individual uncultured organisms is an active area. Methods that use cell sorting to isolate organisms, coupled with sequencing and assembly techniques for single-cell DNA preparations, are producing new genome sequences^{78,79} and, in high-throughput mode, could complement shotgun metagenomics for analysing communities.

One problem associated with genomic data is that it does not address whether an organism is alive or has succumbed to host defences or antibiotic treatment. However, the data can be complemented with transcriptome analysis, or proteomic and metabolomic data sets, which analyse gene expression and metabolic data that are more likely to be derived specifically from living cells.

The simultaneous advances in human genetics and genomics offer opportunities for combining studies of host genotype

with microbiome phenotype. Methods for viewing the microbiome as a quantitative trait and relating this to host genotype are being developed⁸⁰. Advances in host–microbiome studies are also coming from combining immunology and human–microbiome research^{81,82}. Moreover, continued development of statistical methods in microbiome research, such as advances in power analysis, will aid experimental design and future analysis. ■

1. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host–bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
2. Foxman, B., Goldberg, D., Murdock, C., Xi, C. & Gilsdorf, J. R. Conceptualizing human microbiota: from multicelled organ to ecological community. *Interdiscip. Perspect. Infect. Dis.* **2008**, 613979 (2008).
3. Possemiers, S., Bolca, S., Verstraete, W. & Heyerick, A. The intestinal microbiome: a separate organ inside the body with the metabolic potential to influence the bioactivity of botanicals. *Fitoterapia* **82**, 53–66 (2011).
4. Shanahan, F. The host–microbe interface within the gut. *Best Pract. Res. Clin. Gastroenterol.* **16**, 915–931 (2002).
5. Bruls, T. & Weissenbach, J. The human metagenome: our other genome? *Hum. Mol. Genet.* **20**, R142–R148 (2011).
6. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
This paper presents initial findings on the gut microbiome from the MetaHIT project.
7. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).
This paper presents analysis of data from the HMP.
8. Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H. & DeLong, E. F. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**, 591–599 (1996).
9. Vergin, K. L. *et al.* Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Appl. Environ. Microbiol.* **64**, 3075–3078 (1998).
10. Olsen, G. J., Lane, D. J., Giovannoni, S. J., Pace, N. R. & Stahl, D. A. Microbial ecology and evolution: a ribosomal RNA approach. *Annu. Rev. Microbiol.* **40**, 337–365 (1986).
11. Manichanh, C. *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* **55**, 205–211 (2006).
12. Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. Molecular identification of bacteria associated with bacterial vaginosis. *N. Engl. J. Med.* **353**, 1899–1911 (2005).
13. Flint, H. J., Bayer, E. A., Rincon, M. T., Lamed, R. & White, B. A. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nature Rev. Microbiol.* **6**, 121–131 (2008).
14. Srikanth, C. V. & McCormick, B. A. Interactions of the intestinal epithelium with the pathogen and the indigenous microbiota: a three-way crosstalk. *Interdiscip. Perspect. Infect. Dis.* **2008**, 626827 (2008).
15. Jakobsson, H. E. *et al.* Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS ONE* **5**, e9836 (2010).
16. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol.* **6**, e280 (2008).
17. Miller, C. P., Bohnhoff, M. & Rifkind, D. The effect of an antibiotic on the susceptibility of the mouse's intestinal tract to *Salmonella* infection. *Trans. Am. Clin. Climatol. Assoc.* **68**, 51–55 (1956).
18. Sekirov, I. *et al.* Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infect. Immun.* **76**, 4726–4736 (2008).
19. Crowell, A., Amir, E., Tegatz, P., Barman, M. & Salzman, N. H. Prolonged impact of antibiotics on intestinal microbial ecology and susceptibility to enteric *Salmonella* infection. *Infect. Immun.* **77**, 2741–2753 (2009).
20. Mulligan, M. E. Epidemiology of *Clostridium difficile*-induced intestinal disease. *Rev. Infect. Dis.* **6**, S222–S228 (1984).
21. Jarchum, I. & Pamer, E. G. Regulation of innate and adaptive immunity by the commensal microbiota. *Curr. Opin. Immunol.* **23**, 353–360 (2011).
22. Marsland, B. J. Regulation of inflammatory responses by the commensal microbiota. *Thorax* **67**, 93–94 (2012).
23. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
24. Gough, E., Shaikh, H. & Manges, A. R. Systematic review of intestinal microbiota transplantation (fecal bacteriotherapy) for recurrent *Clostridium difficile* infection. *Clin. Infect. Dis.* **53**, 994–1002 (2011).
25. Brandt, L. J. & Reddy, S. S. Fecal microbiota transplantation for recurrent *clostridium difficile* infection. *J. Clin. Gastroenterol.* **45**, S159–S167 (2011).
26. D'Onofrio, A. *et al.* Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem. Biol.* **17**, 254–264 (2010).
27. Human Microbiome Jumpstart Reference Strains Consortium. A catalog of reference genomes from the human microbiome. *Science* **328**, 994–999 (2010).
This paper presents methods and analysis for large-scale production of reference genome sequences from human-microbiome organisms.
28. Gans, J., Wolinsky, M. & Dunbar, J. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science* **309**, 1387–1390 (2005).
29. Nelson, T. A. *et al.* PhyloChip microarray analysis reveals altered gastrointestinal microbial communities in a rat model of colonic hypersensitivity. *Neurogastroenterol. Motil.* **23**, 169–177 (2011).
30. Bent, S. J. *et al.* Measuring species richness based on microbial community fingerprints: the emperor has no clothes. *Appl. Environ. Microbiol.* **73**, 2399–2401 (2007).
31. Sogin, M. L. *et al.* Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl Acad. Sci. USA* **103**, 12115–12120 (2006).
32. The NIH HMP Working Group *et al.* The NIH Human Microbiome Project. *Genome Res.* **19**, 2317–2323 (2009).
33. Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* **486**, 215–221 (2012).
This paper describes the data sets and resources of the HMP.
34. Lazarevic, V. *et al.* Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J. Microbiol. Methods* **79**, 266–271 (2009).
35. Claesson, M. J. *et al.* Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res.* **38**, e200 (2010).
36. Gloor, G. B. *et al.* Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS ONE* **5**, e15406 (2010).
37. Hummelen, R. *et al.* Deep sequencing of the vaginal microbiota of women with HIV. *PLoS ONE* **5**, e12078 (2010).
38. Zubrzycki, L. & Spaulding, E. H. Studies on the stability of the normal human fecal flora. *J. Bacteriol.* **83**, 968–974 (1962).
39. Luckey, T. D. Introduction to intestinal microecology. *Am. J. Clin. Nutr.* **25**, 1292–1294 (1972).
40. Arumugam, M. *et al.* Enterotypes of the human gut microbiome. *Nature* **473**, 174–180 (2011).
41. Martin, J. *et al.* Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. *PLoS ONE* **7**, e36427 (2012).
42. Wu, G. D. *et al.* Linking long-term dietary patterns with gut microbial enterotypes. *Science* **334**, 105–108 (2011).
43. Ravel, J. *et al.* Vaginal microbiome of reproductive-age women. *Proc. Natl Acad. Sci. USA* **108**, S4680–S4687 (2011).
44. Proctor, L. M. The Human Microbiome Project in 2011 and beyond. *Cell Host Microbe* **10**, 287–291 (2011).
45. DOE Joint Genome Institute. A *Genomic Encyclopedia of Bacteria and Archaea*. <http://www.jgi.doe.gov/programs/GEBA/> (US Department of Energy, 2012).
46. Parfrey, L. W., Walters, W. A. & Knight, R. Microbial eukaryotes in the human microbiome: ecology, evolution, and future directions. *Front. Microbiol.* **2**, 153 (2011).
47. Wylie, K. M., Weinstock, G. M. & Storch, G. A. Emerging view of the human virome. *Transl. Res.* <http://dx.doi.org/10.1016/j.trsl.2012.03.006> (24 April 2012).
48. Breitbart, M. *et al.* Metagenomic analyses of an uncultured viral community from human feces. *J. Bacteriol.* **185**, 6220–6223 (2003).
49. Palacios, G. *et al.* Panmicrobial oligonucleotide array for diagnosis of infectious diseases. *Emerg. Infect. Dis.* **13**, 73–81 (2007).
50. Wang, D. *et al.* Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol.* **1**, E2 (2003).
51. Casas, V. & Rohwer, F. Phage metagenomics. *Methods Enzymol.* **421**, 259–268 (2007).
52. Allander, T. *et al.* Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc. Natl Acad. Sci. USA* **102**, 12891–12896 (2005).
53. Finkbeiner, S. R. *et al.* Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathogens* **4**, e1000011 (2008).
54. Breitbart, M. & Rohwer, F. Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques* **39**, 729–736 (2005).
55. Wylie, K. M., Mihindukulasuriya, K. A., Sodergren, E., Weinstock, G. M. & Storch, G. A. Sequence analysis of the human virome in febrile and afebrile children. *PLoS ONE* **7**, e27735 (2012).
56. Pride, D. T. *et al.* Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *ISME J.* **6**, 915–926 (2011).
57. Minot, S., Grunberg, S., Wu, G. D., Lewis, J. D. & Bushman, F. D. Hypervariable loci in the human gut virome. *Proc. Natl Acad. Sci. USA* **109**, 3962–3966 (2012).
58. Breitbart, M. *et al.* Viral diversity and dynamics in an infant gut. *Res. Microbiol.* **159**, 367–373 (2008).
59. Minot, S. *et al.* The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res.* **21**, 1616–1625 (2011).
60. Willner, D. & Furlan, M. Deciphering the role of phage in the cystic fibrosis airway. *Virulence* **1**, 309–313 (2010).
61. Lepage, P. *et al.* Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* **57**, 424–425 (2008).
62. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* **38**, D355–D360 (2010).

63. Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8**, e1002358 (2012).
64. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).
65. Cantarel, B. L., Lombard, V. & Henrissat, B. Complex carbohydrate utilization by the healthy human microbiome. *PLoS ONE* **7**, e28742 (2012).
66. Turnbaugh, P. J. & Gordon, J. I. The core gut microbiome, energy balance and obesity. *J. Physiol. (Lond.)* **587**, 4153–4158 (2009).
67. Raes, J., Foerstner, K. U. & Bork, P. Get the most out of your metagenome: computational analysis of environmental sequence data. *Curr. Opin. Microbiol.* **10**, 490–498 (2007).
68. Wooley, J. C., Godzik, A. & Friedberg, I. A primer on metagenomics. *PLoS Comput. Biol.* **6**, e1000667 (2010).
69. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
70. Haas, B. J. *et al.* Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **21**, 494–504 (2011).
71. Jumpstart Consortium Human Microbiome Project Data Generation Working Group. Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* **7**, e39315 (2012).
72. Schloss, P. D., Gevers, D. & Westcott, S. L. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* **6**, e27310 (2011).
73. Wright, E. S., Yilmaz, L. S. & Noguera, D. R. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl. Environ. Microbiol.* **78**, 717–725 (2012).
74. Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J. & Knight, R. UniFrac: an effective distance metric for microbial community comparison. *ISME J.* **5**, 169–172 (2011).
75. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
76. Caporaso, J. G. *et al.* Moving pictures of the human microbiome. *Genome Biol.* **12**, R50 (2011).
77. Angiuoli, S. V., White, J. R., Matalaka, M., White, O. & Fricke, W. F. Resources and costs for microbial sequence analysis evaluated using virtual machines and cloud computing. *PLoS ONE* **6**, e26624 (2011).
78. Chitsaz, H. *et al.* Efficient *de novo* assembly of single-cell bacterial genomes from short-read data sets. *Nature Biotechnol.* **29**, 915–921 (2011).
79. Dichosa, A. E. *et al.* Artificial polyploidy improves bacterial single cell genome recovery. *PLoS ONE* **7**, e37387 (2012).
80. Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc. Natl Acad. Sci. USA* **107**, 18933–18938 (2010).
81. Elinav, E. *et al.* NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757 (2011).
82. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).
83. Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J. & Goodman, R. M. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**, R245–R249 (1998).
84. Riesenfeld, C. S., Schloss, P. D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* **38**, 525–552 (2004).
85. Hooper, L. V. & Gordon, J. I. Commensal host–bacterial relationships in the gut. *Science* **292**, 1115–1118 (2001).

Acknowledgements The author gratefully acknowledges generous support from the National Institutes of Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this article at go.nature.com/1oqsjuw. Correspondence should be addressed to G.W. (gweinsto@genome.wustl.edu).